ANALYSIS OF A DECISION:

A critique of the National Marine Fisheries Service's draft Biological Opinion on the Operation of the Federal Columbia River Hydropower System, from the perspective of the sciences of decision analysis and risk assessment

Prepared by

Gretchen R. Oosterhout, Ph.D.
Decision Matrix, Inc.
PO Box 1127
Eagle Point, OR 97524
dmatrix@teleport.com

Prepared for

Columbia River Inter-Tribal Fish Commission
Trout Unlimited
Save Our Wild Salmon
Earthjustice Legal Defense Fund
American Rivers

October 9, 2000

TABLE OF CONTENTS

Acronyms and abbreviations

Acronym	Definition
BRWG	Biological Requirements Work Group.
CRI	Cumulative Risk Initiative.
FCRPS	Federal Columbia River Power System.
NMFS	National Marine Fisheries Service.
NRC	National Research Council.
PATH	Plan for Analyzing and Testing Hypotheses.
RPA	Reasonable and Prudent Alternative.
SRSSC	Snake River spring and summer chinook.
STUFA	State, Tribal, and U.S. Fisheries Agencies.

Acknowledgements

Thanks to those who provided technical assistance, review, and financial support. Dr. Barry Anderson (Portland State University, Emeritus Professor of Psychology), Earl Weber and Rob Lothrop (Columbia River Inter-Tribal Fish Commission), Charlie Petrosky (Idaho Department of Fish and Game), Paul Wilson (U.S. Fish and Wildlife Service), Rob Masonis (American Rivers), and Todd True (Earthjustice Legal Defense Fund) provided technical assistance and review. Thanks also to Dr. David Marmorek (ESSA Technologies, Vancouver, BC) for analytical review. This study was financially supported by Columbia River Inter-Tribal Fish Commission, Save Our Wild Salmon, and Trout Unlimited.

Executive summary

The template for sound decision-making, especially under the Endangered Species Act, is the structured decision-making approach recommended by the National Research Council in its 1995 publication, *Science and the Endangered Species Act*. This kind of decision structuring provides the best available scientific methodology for following the rules of logic to analyze the potential performance of specific alternatives against clear objectives, taking into account the uncertainties, errors, and variabilities of a perhaps-poorly understood past, and an unknown future (Clemen 1996, Keeney 1992, Keeney and Raiffa 1976, NRC 1995, von Winterfeldt and Edwards 1986).

An exhaustive, formal, peer-reviewed decision analysis actually has been carried out (by PATH, the Plan for Analyzing and Testing Hypotheses) for the Snake River spring and summer chinook (SRSSC) index populations. PATH, a collaborative study that involved scientists from many different organizations, is an excellent example of how the best available decision science can be applied to development of, and utilization of, the best available biological science, particularly under uncertainty. In its draft biological opinion for Federal Columbia River Power System (FCRPS) operations released in July, 2000, NMFS elected not to make much use of it. Instead, NMFS offers a collection of models and arguments, which they say is "…a chain of simple analyses" (CRI 2000a, p. 13) that provide a "… chain of arguments in order to simplify the complexity of nature" (CRI 2000a, p. 14).

This paper evaluates the "chain of simple analyses [and] arguments" in the draft Biological Opinion from the perspective of the sciences of decision analysis and risk assessment. The draft Biological Opinion addresses the effects of FCRPS operations on all of the ESA-listed salmon ESUs affected by the FCRPS, but the focus of this paper is on SRSSC. The focus on SRSSC is appropriate because, under the ESA, the draft Biological Opinion must reach a decision that avoids jeopardy for *all* listed salmon ESUs, which means specifically avoiding jeopardy for *each* salmon ESU. If the draft Opinion fails to accomplish this purpose for any ESU(s), its analysis and conclusions, at least for those ESUs, should be revisited. In addition, if an evaluation of data and analysis focused on SRSSC shows that the analysis in the draft Biological Opinion is flawed for this ESU, at least some of the flaws also may apply to the draft Biological Opinion's analysis for other ESUs. As explained in the pages that follow, the most fundamental conclusion about the logic of the analysis described in the draft Biological Opinion is that some of the most important elements of rational decision making are missing or inadequate, and that as a result, the analyses do not support the conclusion drawn.

The template of this paper is textbook decision analysis; but the fundamentals that are emphasized are based on straightforward logic. From this perspective, the most important errors, omissions, and inadequacies in the draft Biological Opinion are:

The "Reasonable and Prudent Alternative" is poorly defined and thus impossible to analyze.

If the draft Biological Opinion were based on an appropriately structured decision as recommended by the NRC, the analyses would be focused on the performance of some clearly defined alternatives. Since only hydrosystem improvements have been defined and analyzed, these analyses only predict that the Reasonable and Prudent Alternative (RPA) can meet the jeopardy standards for 4 of the SRSSC index streams (best case scenarios); or for 3 of the 7 (what NMFS calls "worst case" scenarios)[1]. The required additional improvements in survival due to offsite mitigation are assumed to be provided by the "Performance Standards" process, though the mechanisms have yet to be defined.

Performance standards are important when they clearly define goals in terms of measurable attributes; but performance standards are no more than target attributes; they are not actions. The only thing such performance standard definitions really indicate is what might need to be accomplished, and thus how far short of avoiding jeopardy the proposal actually falls. Many of the actions to satisfy these performance standards have not as yet been defined, and many of them are not under the control of the decision-makers because they are owned or managed by private, state, or other entities. Identifying what needs to be accomplished is very important, but actually defining a particular set of actions and showing that they are likely to be implemented and to meet the standards, is a different story altogether — one that the draft Biological Opinion fails to tell.

A realistic range of potential outcomes has not been considered.

A rational decision process would take into account the sources, and appropriate ranges, of uncertainty, variability, and error. Risk assessment should focus on failure mechanisms, how likely they are to impact potential outcomes, and how to detect, mitigate, and control them. This is standard practice (Burgman et al. 1993, Glickman and Gough 1990, Hilborn and Walters 1992, Vose 2000), but it is not what NMFS has done.

The draft Biological Opinion acknowledges here and there that many of its simplifying assumptions might be wrong, and, to the CRI's credit, some of its assumptions have actually been analyzed. However, when some purportedly "worst case" analyses are considered in the draft Biological Opinion, they are summarily rejected simply because NMFS deems them "beyond reason," regardless of the extensive justification for them developed by PATH or STUFA. The ultimate conclusions (that the RPA and offsite mitigation efforts will avoid jeopardy and promote recovery) are based on analyses that fail to take into account what could go wrong, how likely such problems could be, what their impacts could be, and how such problems could be controlled or mitigated. The discussion in the draft Biological Opinion mentions some of these issues occasionally, but the ultimate conclusions, for the SRSSC and the other ESUs, ignore them.

Moreover, almost every assumption on which the draft Biological Opinion actually relies is optimistic; it is as though the less-appealing branches of a decision tree had been cut away. The models assume that improvements in demographic variables occur immediately and permanently, with no transition period or delay. There is no allowance for demographic, genetic, or environmental variability or uncertainty. There is no allowance for environmental or

[1] This statement is based on the updated draft Biological Opinion's Table 6.3-13 dated Sept. 12, 2000.

demographic trends. The analyses do not distinguish between natural and human-induced effects, and none takes feasibility into account. No evidence is provided that the kinds of survival increases posited in the "numerical experiments" or sensitivity analyses have ever been produced anywhere, particularly so quickly. These flaws in the analysis affect the conclusions of the draft Biological Opinion for all ESUs, not just the SRSSC.

The analyses do not make full use of the best available science for Snake River spring and summer chinook.

Standardizing analysis over 12 ESUs may be an appealing goal to some, but extinction risk is not a relative concept for jeopardy decisions under the ESA. Furthermore, standardization of analysis across all Snake and Columbia River salmon ESUs is impossible if the best available science is to be used for each ESU. The ESA does not allow comparative judgments about actions unless those actions all avoid jeopardy and provide for recovery.

Standardization is the justification offered in the draft Biological Opinion for using the absolute extinction threshold, and for relying on uncalibrated, simplistic, and inconsistent models. Some stocks have received vastly greater levels of study, have longer time series of observations, and therefore permit the development of better analytical tools than others. The fact that inconsistent and inadequate data are all that is available for many populations means that research standards should be raised where needed, not that the analytical standard should be lowered when better data and analyses are available.

The way NMFS has attempted to standardize such a complex, diverse set of analytical requirements is by relaxing the analytical standards so much that most, if not all, data sets are assumed to satisfy them, even when some of the data sets do not, in fact, meet the requirements. The metaphor that may best describe this attempt to treat very different problems and data sets as though they are all the same is the observation that when the only tool you know how to use is a hammer, all problems have to be treated like nails. There is a well-developed, widely accepted toolbox readily available, and a rich supply of data for at least some of the ESUs, but NMFS has chosen not to make full use of it.

As a result largely of assumptions made in the name of "standardization," the draft Biological Opinion seriously underestimates the probability of extinction. It also underestimates the target survival increases that are used to define the performance standards, particularly for ESUs such as the SRSSC that are in accelerating decline. Such analyses may be "standardized," but they provide demonstrably inaccurate answers about the ability of the RPA and offsite mitigation measures to achieve the critical ESA objective of avoiding jeopardy.

The models are inappropriate and misleading.

For SRSSC, the equilibrium population growth rate estimated from the modified Dennis model is not a legitimate representation even of this ESU's historical population dynamics because these populations have been in accelerating decline during the period of analysis. A model that relies on a statistical measure of central tendency, whether it is a mean or a median, does not capture the level of threat that is indicated when the rate of decline of a population has been increasing. Neither are the average long-term population growth rates from the Leslie matrix model, or the simple spreadsheet exponential growth projections, legitimate representations of the SRSSC populations in the near future, in part because they also ignore the accelerating decline of the past 20 years, as well as inescapable fluctuations over the short term.

NMFS does not have an analytic tool which can be used to make probabilistic estimates about different management scenarios, only some simple spreadsheet models and Leslie matrix models that can make *deterministic* predictions about long-term equilibria of different scenarios, and another model — the modified Dennis model — that can make *probabilistic* projections of the past. These models are based on different assumptions, though they share some structural similarities, and they produce distinctly different results when they should produce equivalent results; yet NMFS treats them as interchangeable. While it might be possible to conclude that there is less than a 5% probability of extinction associated with a stochastic model's median growth rate of, say, 0.9, based on historical data, that is not the same as being able to say that an equilibrium constant growth rate of 0.9, calculated by a deterministic spreadsheet algebraic model, or a Leslie matrix model, represents a 5% probability of extinction. Yet that is what the CRI does.

Because these various models and analytical assumptions have not been validated against empirical data, and are not well correlated with one another, they produce results that are inconsistent with one another as well as with the available data. For example, Figure 1 shows different expected trajectories calculated by the CRI's extinction model, the management model (Leslie matrix, calibrated to spawner-recruit data), and the spreadsheet models of the RPA with offsite mitigation.

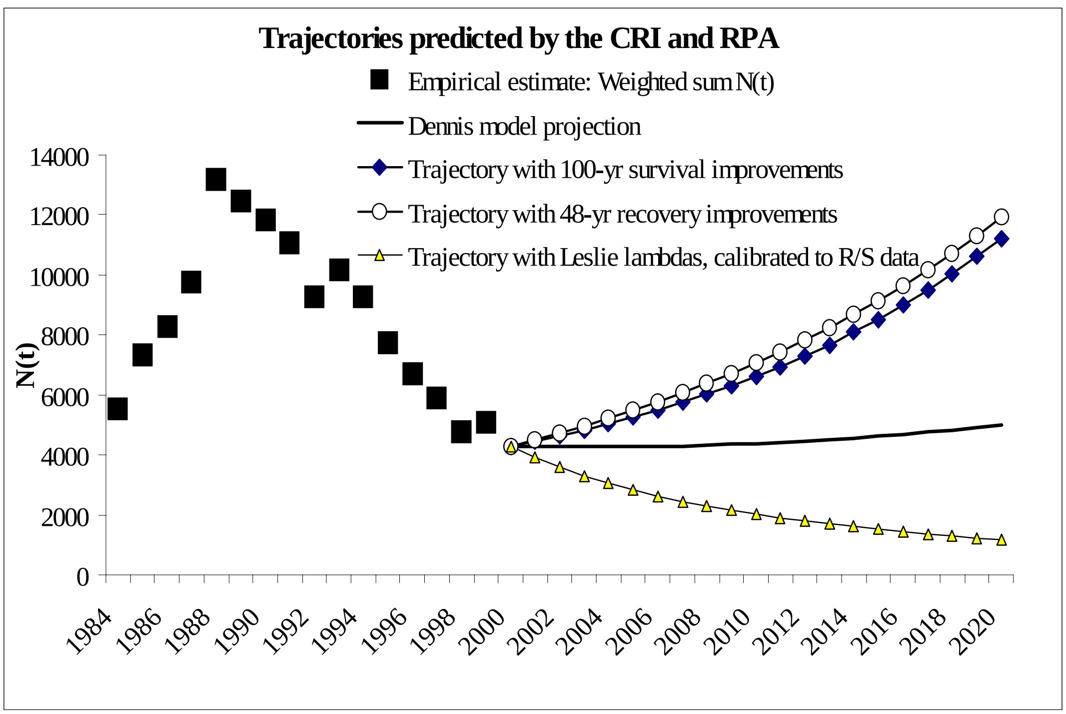

Figure 1. Trajectories predicted by the Dennis model, by the RPA + offsite mitigation survival and 48-year recovery improvements, and by the Leslie matrix, for the aggregate

SRSSC population. Data source: 9/5/00 CRI revisions, Appendix B; and 9/7/00 CRI Excel spreadsheets.

A casual inspection of the different trajectories shown in Figure 1 indicates that the Leslie matrix growth rates (which have been calibrated against spawner-recruit data) most reasonably represent an extrapolation of the current trend; yet the Leslie matrix growth rates are not what are being used in the draft Biological Opinion. Instead, the CRI assumes that the nearly flat trajectory labeled "Dennis model projection" is the best extrapolation of the status quo. This key assumption is not supported by common sense: because the nearly flat trajectory assumed in the Dennis model represents an assumption that the current population is both stable and approximately sustainable, the risk of extinction appears to be very small over 100 years, even with little effort put into recovery — a conclusion simply not supported by the past 20 years of data. Furthermore, because this baseline is based on such an optimistic projection, the target survival improvements are also quite low. If the required increases were based on the most logical extrapolation of the status quo, it is clear that survival increases would need to be significantly greater than NMFS' current assumptions suggest.

The analyses do not appropriately address the metrics.

The analyses of future actions provided in the draft Biological Opinion do not actually estimate the probabilities that these species will avoid absolute extinction and achieve recovery for any of the ESUs, though the draft Opinion promotes the impression that they do. The analyses of future actions are all carried out using simple, deterministic analyses, which cannot provide results in probabilistic terms. The metrics are probabilistic, but the models used for assessing them (Leslie matrix and additive spreadsheet survival calculations) are not.

To be sure, the CRI's Dennis model for estimating probability of extinction is a quasi-stochastic regression forecast, which can, in theory, provide probability estimates under a restrictive set of assumptions. But it cannot properly be applied to analysis of significantly different scenarios because a continuous regression forecasting model can only extrapolate a continuation of past trends. It cannot be used to project the future effects of changed directions because such regressions are continuous; they do not have upward kinks in the middle.

Because the Dennis model cannot appropriately be used for prospective analyses under changing conditions, NMFS has to switch models to evaluate the RPA. In place of an analysis that would directly address the probability of survival or recovery for alternatives such as the RPA or the "proposed action," NMFS uses deterministic exponential growth models, including deterministic Leslie matrix models. The appeal of these simple models may be a legitimate justification if that is the best one can do and if they can be shown to be representative. But these models have not been shown to be representative, even where data are readily available; and considering that stochastic models are readily available which have been offered to NMFS, and that probability distributions are also readily available (from PATH and STUFA[2]) at least for SRSSC index

[2] State, Tribal and U.S. Federal Fishery Agencies, a group of biologists and analysts from USFWS, IDFG, ODFW, CBFWA, and CRITFC.

populations, there is no justification for using such simplistic, deterministic, exponential population growth models to support such an important decision.

The survival metric is not adequate to represent the survival objective.

The absolute extinction survival metric falls short of addressing the fundamental objective of the analysis — to avoid jeopardy — because the absolute extinction standard is arguably the least conservative standard possible, and does not come close to indicating whether a population has a "high" probability of survival. The way NMFS has defined this metric means that, unless a population is forecast to drop to 1 or 0 spawners for 5 years in a row, there is "…a high likelihood that the species' population will remain above critical escapement thresholds over a sufficiently long period of time" (NMFS 2000 p. 1-7). "Standardization" is the only rationale offered by NMFS for choosing absolute extinction as a metric, due to the assertion that adequate data for a more refined metric and analysis are not available for all ESUs and that populations are all different. But the goal of standardization does not justify a failure to make use of the best possible data and analyses that are available for some ESUs (particularly the SRSSC) in order to define a more conservative analytical threshold where possible.

In addition, apart from the availability of data, absolute extinction is a metric or threshold that is rarely, if ever, used in conservation biology because it does not include considerations of depensation, or of genetic, environmental, demographic, or catastrophic uncertainty, and it allows no margin for error. Because no standard threshold has been established for extinction analyses, the ISAB (1999) recommended that NMFS report results using a range of commonly used thresholds; but since that ISAB review, NMFS has actually lowered the threshold from one spawner/one year, to one spawner/one generation. A decision based on this least conservative threshold would be classified as strongly risk-prone, rather than risk-averse, which is more the norm in conservation biology, as it is under the ESA (Noss et al. 1997, NRC 1995). This concern with NMFS' reliance on an absolute extinction threshold applies across the salmon ESUs addressed in the draft Biological Opinion.

The 5- and 8-year checkpoint strategy is an inadequate approach to risk management.

The RPA is in large part a plan to develop plans to detect further degradation, with the assumption that if needed, time will be adequate to find ways to control for the risks that materialize. This approach assumes that (1) not enough is known now to be able to decide to take substantive action, (2) enough will be known in 5-8 years, and (3) actions to control or mitigate further declines could be successfully implemented if needed. Ironically, this is almost exactly the position stated in the 1995 Biological Opinion, which asserted that enough information would be gathered in 4 years to make a drawdown decision in 1999.

The overwhelming scientific judgment outside of NMFS, and the result from the only peer-reviewed, formal decision analysis that has been conducted on SRSSC, is that immediate action that includes bypassing the four lower Snake River dams is critical if the SRSSC ESU is to have adequate prospects of survival, let alone recovery. The approach of setting checkpoints at the 5-year and 8-year point is an appealing start towards risk management, but because a Snake River spring/summer chinook salmon generation is approximately 5 years, it is unlikely that changes in key uncertainties could be detected at these checkpoints, let alone detected clearly enough that some kind of corrective action could be identified, assessed for its efficacy, and implemented. Given the accelerating decline of SRRSC and other Columbia River populations, putting off a decision on dam removal or other major actions for 5 or 8 years is unjustified.

In summary, the draft Biological Opinion is technically deficient in several critical respects. Most notably, it fails to follow sound, well-established principles for making decisions under conditions of uncertainty. The flaws in the draft Biological Opinion serve to underestimate the extinction risk of SRRSC and other populations undergoing accelerating decline, and as a result, to underestimate the magnitude of the population growth changes needed to avert extinction and achieve recovery.

I. Introduction

The purpose of this paper is to evaluate the way the National Marine Fisheries Service (NMFS) is making the decisions set forth in its draft Biological Opinion for the Federal Columbia River Power System (FCRPS) operations (July 27, 2000 draft), specifically about managing Columbia and Snake River salmon populations, in light of their ESA status. The National Research Council (NRC) recommends using decision structuring methods to make decisions about species that fall under the Endangered Species Act (ESA) because decision structuring tools can greatly improve the quality of decisions like these: decisions involving multiple stakeholders, conflicting objectives, and uncertainty (NRC 1995). The science of decision analysis thus provides the appropriate framework from which to evaluate this NMFS decision process.

A variety of approaches can be called "decision structuring," but they all rely on following the rules of logic to structure a decision and analyze it rationally. Decision science evolved from the fields of statistics, cognitive psychology, and optimization theory. From *statistics* come the most reliable ways of analyzing uncertainty and using probabilistic data. From *cognitive psychology* come ways of managing cognitive biases, focusing decisions on objectives, developing utility functions that reflect decision-makers' values and risk aversion, and eliciting expert judgments. From *optimization* come ways of evaluating tradeoffs and uncertainties in order to satisfy objectives.

Structuring a decision requires explicitly defining objectives, developing representative metrics to assess current conditions and measure progress, quantifying uncertainties, and clarifying sources and magnitudes of risk. Decision structuring provides a statistically and psychologically valid framework within which to evaluate management actions. It provides the best available scientific methodology for making a decision, with full consideration of objectives, uncertainty, and risk (Clemen 1996, Keeney 1992, Keeney and Raiffa 1976, NRC 1995, von Winterfeldt and Edwards 1986).

Although decision structuring is solidly founded in these academic fields, it can be as simple as a 2 X 2 decision table, or as complex as a multi-attribute decision tree incorporating Markov/Monte Carlo and Bayesian methods for dealing with uncertainty. Fundamentally, structuring a decision so it best fulfills the dictates of logic and rationality is a matter of following some common-sense rules. Any rational process ought to be clear about what the decision is, be clear about what the objectives are, and make sure the facts and analyses support the conclusions. A rational decision process focuses on what it is possible to do and what it is best to do, what can be controlled and what cannot (NRC 1995, von Winterfeldt and Edwards 1986).

Decision structuring methods include a variety of tools for examining sources of error, environmental variability, and uncertainty. Those tools include sensitivity analysis, stochastic modeling, experimental design, judgment elicitation methods, and Value of Information assessments. The science of risk assessment offers companion tools for analyzing and managing risk. Together with decision structuring, risk assessment helps decision-makers identify ways that uncertainty, variability and error can be analyzed, monitored, mitigated, and controlled.

NMFS has chosen not to use most of these tools ordinarily associated with decision structuring and risk assessment. Consequently, the decisions described in the draft Biological Opinion largely do not follow rationally and logically from the available evidence.

II. Why the NRC calls for decision structuring in making decisions under the ESA

Delaying a decision in order to gather more information is often an appealing choice for politically risk-averse decision-makers. No matter how much more data and research results become available, there is never enough to satisfy everyone; this is because people's preferences tend to be driven by their own values, including their attitudes towards risk, more than by sometimes-ambiguous facts. As a result, the constant search for greater certainty by default supports the status quo (Keeney 1992, NRC 1995, von Winterfeldt and Edwards 1986).

The decision NMFS has proposed in the draft Biological Opinion for Snake River salmon populations provides a good example of why the NRC recommends using decision structuring for ESA decision-making, in part to avoid temptation of seeking unattainable certainty while continuing status quo operations. As the NRC (1995) explains:

> *Those implementing the [Endangered Species] act almost always believe that with additional information, they could make a better decision. Nevertheless, decisions to delay action pending further information…should be viewed critically. What kind of information and of what quality could be gathered within the time and resources available? What are the possible answers that such investigation might reveal? What decisions would be triggered by different answers? How are those decisions different from those that would be made using existing information? What effect will continuing the status quo have on species status and on options for future action? Considering these questions in a structured framework can make it more likely that a reasonable decision will be made (p. 160).*

The NRC's advocacy of decision structuring tools for implementing the ESA is not a call for more extensive analysis or research. A good decision structure can be relatively qualitative, and in the long run should facilitate reaching agreement and producing a decision that is socially and scientifically defensible. As the NRC notes, many managers are uncomfortable with probabilistic reasoning and resort to simplistic or ad-hoc heuristics in order to simplify information and justify their choices. It is this tendency, which often leads to poor decisions, that the NRC seeks to avoid through application of the tools of decision structuring. Decision structuring helps decision-makers avoid a number of common and critical mistakes, by providing tools that help them (NRC 1995):

- **Explicitly identify and prioritize objectives**

  In theory, the ESA provides a ready-made objective: avoid "jeopardy" to a listed species, based on biological considerations alone; if more than one alternative meets that criterion, then other objectives (such as economics) may be considered. Making *all* objectives explicit is fundamentally important to sound decision-making because the objectives define the decision context, regardless of whether they are stated or unstated. In addition, the analysis of the alternatives must be conducted in such a way that the analyses directly relate to the objectives.

For example, PATH was structured as a formal decision analysis and as such, carried out the analyses within a framework that was meticulously defined in terms of the objectives. In contrast, although NMFS does acknowledge the 1995 jeopardy standard as the overriding objective, the analyses in the draft Biological Opinion consist of a set of models and assumptions that only address the jeopardy standard indirectly. As this paper will show, there are several critical mismatches between what NMFS says its analyses show, what those analyses actually show, and how they relate to the jeopardy standard NMFS has articulated as its overriding objective.

- **Establish criteria for measuring or estimating progress**

Because fundamental objectives are rarely directly measurable, decision structuring offers tools to help establish metrics to represent the objectives and facilitate the analysis. Such metrics are useful not only for evaluating the relative potential performance of the alternatives in the decision-making process, but also for using adaptive management feedback mechanisms to incorporate new information about critical uncertainties into future decisions. A thorough analysis of adaptive management should examine the amount of learning that is feasible within various time periods, given uncertainties in system response to actions, environmental variation, and measurement error (Peters et al. 2000). NMFS has not done this.

If metrics are not directly related to achievement of objectives, they can lead to misplaced efforts, wasted resources, and mistaken choices. The decision criteria established by NMFS, in the form of models and still-ambiguous performance standards, fail to meet the requirements of sound decision analysis on several counts, but perhaps most critically because neither the decision criteria nor the performance standards relate directly or clearly to the jeopardy standards.

Setting — and measuring progress towards — performance standards is obviously important. A coherent decision structure provides a useful mechanism for ensuring that performance standards are appropriate to the objectives. Performance standards can also provide feedback mechanisms which allow decision-makers to incorporate new information into corrective responses. But unless the actions intended to meet the stated performance standards are clearly identified, appropriate to the objectives, feasible, funded, and committed to, then simply identifying some performance standards is not likely to contribute to achieving objectives.

- **Weigh tradeoffs**

It is not too unusual, in a decision where the best science may not be compatible with some short-term economic or political interests, for there to be unstated objectives which play an important role in the decision instead of, or in addition to, the legally mandated or explicitly stated objectives. As the NRC (1995) puts it,

> *The [ESA] act prohibits consideration of human objectives unrelated to species*

> *protection in decisions regarding listing, take, and jeopardy, but directs that these other objectives be taken into account in decisions about critical habitat and implementation of recovery plans. Tradeoffs between species protection and economic or other benefits or costs must [then] be evaluated. Again, because these tradeoff decisions are often difficult and controversial, it is important to use well-structured and explicit methods for making them (p. 14).*

In a structured decision process, such unstated tradeoffs are forced to the surface, illuminated and made explicit so that the mechanisms truly driving a decision are plain. There is little evidence that NMFS plans to evaluate such tradeoffs within a structured decision process, especially for recovery planning.

Risk aversion is one of the most important aspects of how people trade off objectives, and can be a powerful influence on the existence of hidden objectives (Howard 1989, Keeney 1992, Keeney and Raiffa 1976). Tradeoffs are not allowed for decisions regarding jeopardy, but since there is no mention of making tradeoffs between the jeopardy objectives and any non-biological objectives that may underlie the draft Biological Opinion, NMFS should be following the prevailing practice in conservation biology, which is to be risk averse for decisions affecting species extinction (Noss et al. 1997). Yet the agency's analyses in a number of critical areas are distinctly risk prone, suggesting unstated objectives or unanalyzed tradeoffs.

- **Efficiently use available information**

In evaluating the likelihood of meeting its performance standards and achieving its objectives, NMFS makes selective use of available information. Consequently, it appears to ignore some relevant information and use inappropriate species and geographic areas in other parts of its analysis (Oosterhout et al. 2000). Decision structuring does not provide a guarantee against such errors, but if all the pieces and relationships in a decision are clearly defined, it is less likely that important information will fall through the cracks or be used in a misleading way.

- **Explicitly consider uncertainties and probabilistic information**

The methods used by NMFS in the draft Biological Opinion and its supporting analyses to assess the alternatives assume that environmental and demographic equilibrium is immediately achieved, with certainty, and with no environmental or other sources of variability. Analysis of the alternatives themselves is deterministic, with no environmental, demographic, genetic, or catastrophic variability. Implementation and system response are modeled with complete confidence. Median estimates are treated as though they are certain, when by definition, a median is only the 50-50 point in a statistical distribution. Probabilistic information is not incorporated into the analysis.

The Dennis extinction modeling of the 1980-1999 period does estimate some uncertainty in a general sense (as a "spread" around a regression based on historical spawner-recruit data), but in analyses of the RPA, potential impacts of adverse events — even events that seem particularly likely, such as droughts and floods that occur on a regular cycle in the region —

are not adequately taken into account.

- **Explicitly incorporate an assessment of risk aversion into the analysis**

The draft Biological Opinion's pattern of optimistic assumptions and careless accounting for problems that are likely to occur in implementation results not in a risk *averse* approach as required under the ESA, but instead in an approach that is ecologically risk *prone*. The NRC (1995) advocates taking an approach that is more consistent with the "precautionary principle," and doing so within a structured decision framework so that the consequences of being too risk averse, as well as being not risk averse enough, can be explicitly taken into account. (Akcakaya et al. 2000). Indeed, the United States is a signatory to the United Nations agreement on the precautionary approach.

- **Make better use of experts via judgment elicitation**

Decision analysis provides powerful tools for eliciting expert judgments and incorporating them into decisions, particularly through the use of statistical tools, model validation techniques, and expert judgment elicitation. PATH, for example, carried out formal Weight of Evidence judgment elicitations conducted by professional decision analysts. Considering the extent to which NMFS has emphasized the inadequacy of available data, its failure to use any of the available judgment elicitation tools, either those that were used in PATH or others, should be explained.

Research shows that "in-house" self-reinforcing approaches for reaching judgments often lead to poor decision-making. In a landmark study of the phenomenon, Irving Janis christened the problem "Groupthink," and showed that making poor judgments and thus poor decisions is particularly likely when small groups work under internal or external political pressure (Janis 1972). Problems associated with what Janis called "Groupthink," which have been extensively studied and documented, are for the most part avoidable, if more objective and explicit expert elicitation methods are used (Keeney 1992, Russo and Shoemaker 1989, von Winterfeldt and Edwards 1986), and the process makes good use of external review. NMFS chose to extract itself from a multi-agency collaborative process (PATH) that included formal judgment elicitation methods as well as a formal peer review process. Instead NMFS conducted its subsequent analyses entirely "in-house." The evidence available in the draft Biological Opinion indicates that NMFS has failed to avoid these problems.

Most important, decision structuring helps decision-makers ensure that all the pieces of the decision are well developed and coherently linked together, and thus, that the facts and analysis support the conclusions. Ultimately, it is the failure to achieve this congruence between facts, analysis, and conclusions that is most striking about the decision process evidenced in the draft Biological Opinion.

III. How they got here

NMFS has been through several efforts to analyze the effects of FCRPS operations on listed salmon since it issued its 1995 Biological Opinion. The two that are of interest here are the Plan for Analyzing and Testing Hypotheses (PATH) which proceeded from 1995 through about the end of 1999 and the Cumulative Risk Initiative (CRI) which began in early 1999 and provides most of the analysis for the draft Biological Opinion.

III.1. PATH

PATH was a formal biological decision analysis designed to quantitatively evaluate different actions for future operation of the hydrosystem. PATH used criteria defined explicitly in terms of the objectives for survival and recovery of listed Snake River salmon and steelhead. The decision analysis explicitly analyzed a wide range of uncertainties about past stock performance, effectiveness of management actions, and future climate, in the evaluation of each of several detailed hydrosystem management strategies.

In PATH, the decision analysis had four components:

1) retrospective analysis: evaluation of different assumptions (hypotheses) about how environmental factors affected past survival of salmon and steelhead;

2) prospective analysis: analysis of how each hydrosystem management strategy might play out in the future, under different combinations of assumptions about past stock performance, effectiveness of management actions, and future climate;

3) risk analysis: estimation of the likelihood of meeting the ESA standards for survival and recovery of listed salmon and steelhead under each hydrosystem action, based on projections of the number of spawning salmon and steelhead over a range of assumptions and hypotheses;

4) weight of evidence analysis: an assessment of the likelihood that certain key assumptions in the retrospective, prospective, and risk analyses are true based on a comprehensive evaluation of evidence for and against each assumption.

PATH also conducted a wide variety of sophisticated sensitivity analyses, including the possible effects of a range of assumptions about habitat, harvest, hydro, and hatchery actions. These sensitivity analyses were used to evaluate the relative impact of different uncertainties on the rank ordering of the alternatives and outcomes. The key findings for SRSSC (as well as Snake River Fall Chinook) were that alternatives which included breaching the four lower Snake River dams had the highest probability of meeting the jeopardy standards set by NMFS, that no other alternatives had a significant likelihood of meeting this standard, and that the uncertainties that had significant impact on the rank-ordering of the alternatives had to do with delayed mortality. PATH also found that it would take decades even to begin to resolve these uncertainties, that any investigation of delayed mortality would have to use hatchery fish because there are not enough wild fish left, and that the most likely findings would not change the conclusions (Peters et al. 2000).

III.2. Cumulative Risk Initiative (CRI)

NMFS apparently did not find the results and conclusions from PATH acceptable, and so, a few months before the "1999" decision was due, NMFS initiated the Cumulative Risk Initiative (CRI). Although much of the attention focused on the CRI has centered around the Dennis model extinction analyses and the Leslie matrix life cycle models, the CRI encompasses a number of other efforts. Rather than producing "a single decision-theoretic model," NMFS promised that the CRI would offer "a wide variety of scientific dialogues and diverse approaches," along with "…a commitment to transparency and data access" (Kareiva et al. 1999 (draft), p. 5).

The CRI was initiated as an effort "…to coordinate research efforts within the Northwest Fisheries Science Center (NWFSC)" (Kareiva et al. 1999 (draft), p. 4), with stated goals of integrating risk factors, developing definitions for population viability, and identifying productive habitat. A primary goal was to organize data and "provide an analytical framework for integrating the effects of taking actions in different portions of fish life cycles" (Kareiva et al. 1999 (draft), p. 1). Although an early draft report that described what the CRI's approach would be was titled "An introduction to NMFS decision-support science for ESA decision making, with examples," the focus was (and has continued to be) on modeling rather than decision analysis or decision structuring. The CRI initially identified four facets to its approach:

1. Data exploration (evidence that habitat, harvest, hatcheries and hydropower function as risks for recovery, identifying and testing hypotheses);

2. Identification of key risk factors for each population or ESU (risk analysis and sensitivity analysis to determine how alterations in each of the different H's affect extinction risk or population recovery probabilities);

3. Detailed analysis and evaluation of management options (more detailed modeling and research to determine what evidence there is that population trends can be affected, which management options will produce the best results, what the uncertainties are, and research needs);

4. Adaptive management (ecological experiments as conservation actions are implemented).

The CRI's primary analytical tools are the Dennis extinction model and the Leslie matrix life cycle models. These analyses have been used primarily for facet number 2. The draft Biological Opinion and All-H papers describe plans to develop programs to conduct hypothesis testing (facet number 1), detailed evaluation of management options (facet No. 3), and adaptive management (facet No. 4) (CRI 1999, CRI 2000b). To date, however, the CRI analysis has not addressed the steps it originally committed to, let alone the sequence of steps involved in a structured and rational decision.

IV. Standardization

One of the key goals that NMFS has emphasized in the CRI process is "standardization." The agency has argued that it is an important, even an overarching goal, to standardize the risk assessments across all 12 of the salmonid ESUs listed in the Snake/Columbia River basin (CRI 2000a, CRI 2000b).

The ESA, however, does not allow comparative judgments about actions for different species or ESUs that do not also, at a minimum, avoid jeopardy and provide for recovery. In other words, standardization may be an acceptable analytic objective so long as it does not interfere with the substantive objective of avoiding actions that would jeopardize a species or ESU. When populations are at such unquestionably high levels of risk as Snake River spring and summer chinook, the desire for analytic standardization across ESUs cannot be allowed to mask or obscure the quality of the decision for an individual ESU. Standardization is useful only up to a point: it must give way where it undermines the quality of the risk assessment and decision analysis for a particular ESU. Put another way, the objective set by the ESA is to avoid jeopardy, not to achieve standardized risk assessments or decision-making.

If one examines NMFS emphasis on standardization from the perspective of decision analysis under the ESA, two fundamental aspects of it that are problematic:

IV.1. Extinction is not relative.

Under the law, the question about what to do for a listed species is not a relative question. The statute does not ask about what the rank ordering is of populations that are at risk. Even if there were a tool that could be used to calculate risks of extinction accurately and precisely enough to compare the probability of losing one population to the probability of losing another, the law does not authorize NMFS to make decisions about the jeopardy risk of proposed actions based on such a ranking. Instead, it requires decisions that avoid jeopardy for each species regardless of the relative risk among species.

At the same time, so long as biological opinions NMFS issues avoid jeopardy to each species, NMFS may make relative choices about where to focus its efforts. The ability to make these choices, however, cannot lead the agency to adopt, in the name of standardization, an analysis that leads to incorrect or implausible conclusions about whether a proposed action or set of actions will avoid jeopardy. In an over-emphasis on standardization in order to allow comparisons among actions for different ESUs, NMFS has developed analytic tools that discount or ignore evidence relevant to the crucial issue of determining whether proposed actions jeopardize some ESUs.

It may be that NMFS has erred because it is confusing the ranking of risks, with the ranking of where resources should best be invested. Prioritizing actions simply by ranking risks is not a rational way of making choices (Howard 1968, Howard 1989, Keeney 1992, Keeney and Raiffa 1976, von Winterfeldt and Edwards 1986) because it does not take into account feasibility and issues of optimization. In decision science, risk enters into the analysis two ways: via the uncertainties (probabilities of uncertain events and achieving the various possible outcomes), and via the risk aversion reflected in the values and preferences of the decision-makers. Sacrificing analytical rigor in the name of a potentially misleading and inappropriate goal like "standardization" distracts decision-makers from the job that they really need to do: decide how best to save each of the populations that is listed under the ESA, and how to prevent other populations from falling under the ESA.

IV.2. Standardization is impossible.

There is one other problem with the CRI's efforts to standardize risk analyses: unless the ESUs are also standardized, standardizing the analysis in a way that makes adequate use of the available data is impossible.

Watersheds and ecosystems in the Columbia River Basin come in different sizes, ecological diversity, climate regimes, productivity, hatchery influence, degree of urbanization, and the availability and quality of the data. The way NMFS has attempted to standardize such a complex, diverse set of analytical requirements is by relaxing the analytical standards so much that all or most all data sets are assumed to satisfy them, even when some of the data sets do not, in fact, meet the requirements. The metaphor that may best describe this attempt to treat very different problems and data sets as though they are all the same is the observation that when the only tool you know how to use is a hammer, all problems have to be treated like nails. There is a well-developed, widely accepted toolbox readily available, and a rich supply of data for at least some of the ESUs, but NMFS has chosen not to make full use of it.

One example of how NMFS' focus on standardization distorts its reasoning is that the agency applies the Dennis extinction model to populations that are in accelerating decline, but treats them as though they are stable in an attempt to satisfy the conditions of the model. This means that the risk of extinction for the most seriously at-risk populations is, as the CRI acknowledges, probably under-estimated; but under-estimated or not, these risk estimates are compared to estimates for populations which the available data indicates may actually be stable.

In short, applying the CRI models (Dennis extinction model, Leslie matrix models, and spreadsheet analyses) across the board to diverse populations in different ecological conditions, regardless of whether the requirements of the models are satisfied for the populations, may be "standardization," but it is not a defensible way to make decisions about individual threatened and endangered species or ESUs. The analysis may be "standardized," but it provides inaccurate answers about the ability to achieve the critical statutory objective of avoiding jeopardy. It would be more appropriate to acknowledge that the risk of extinction is high for all of the salmonid ESUs, and then use the best available science and analytical tools, including decision structuring tools, as recommended by the NRC (1995).

V. Objectives

The most critical element of most decisions is a clear definition of the objectives, because objectives provide the framework by which the alternatives are evaluated (Derby and Keeney 1990, Keeney 1992, Keeney and Raiffa 1976, von Winterfeldt and Edwards 1986). As the NRC puts it, "A hallmark of formal decision analysis…and other structured problem-solving methods…is an emphasis on articulating clearly the objectives for a decision and criteria for evaluating how well alternative proposals might meet those objectives" (NRC 1995, p. 159).

The rationality of a decision is determined by how well the analysis underlying the decision supports the conclusions, as measured against the identified objectives (von Winterfeldt and Edwards 1986). If the objectives are not clear, or the analyses do not address the potential performance of the alternatives in terms of those objectives, then the logic and rationality of a decision are flawed. If the logic and rationality of the decision are flawed, then the outcome is less likely to be one that meets the actual objectives. The more important the decision is, the more important it is to be sure the logic of the analysis actually does support the conclusion drawn — and that the conclusion drawn supports attainment of the objectives.

Under the ESA, the criterion for choosing among alternatives is, in theory at least, quite clear: an alternative must be found that will avoid jeopardy to listed species. For the actions and alternatives at issue here (i.e., operation of the FCRPS and its effects on SRSSC), NMFS continues to cite the 1995 FCRPS Biological Opinion, which defines the objectives as follows (NMFS 2000):

> *At the species level, NMFS considers that the biological requirements for survival, with an adequate potential for recovery, are met when there is a high likelihood that the species' population will remain above critical escapement thresholds over a sufficiently long period of time. Additionally, the species must have a moderate to high likelihood that its population will achieve its recovery level within an adequate period of time. The particular thresholds, recovery levels, and time periods must be selected depending upon the characteristics and circumstances of each salmon species under consultation (p. 1-7).*

If more than one course of action can be identified that will satisfy this biological objective, then economic, social, or other objectives can be considered.

This statement of objectives seems clear enough in the draft Biological Opinion, but the following discussion about metrics shows that the draft Biological Opinion fails to satisfy the NRC's standard for objectives in decision structuring cited above because "…the criteria for evaluating how well alternative proposals might meet those objectives" (NRC 1995, p. 159) do not adequately represent the objectives, and the analyses do not adequately support the conclusions reached.

V.1. Metrics

Objectives themselves are rarely directly measurable. It is usually necessary to define some metrics, or measurable attributes that, if met, indicate the objective will also be met.

Decision analysts have identified key criteria to evaluate how well a set of attributes or metrics will serve as measures for achieving an objective. Using appropriate metrics helps ensure that a decision structure will accurately represent the decision being made and thus have a higher likelihood of producing the desired outcome. Appropriate metrics also help ensure that the decision makes sense to stakeholders (Keeney 1992, Keeney and Raiffa 1976, vonWinterfeld and Edwards 1986). Two key requirements in defining a metric are (1) that the metric, and the method of analyzing it, clearly relate to the objective; and, (2) that the metric itself is feasible to apply.

It is well established that a qualitative descriptor like "high likelihood" means different things to different people, and that what it means depends on how it is framed, particularly whether it relates to a potential loss or a gain, and whether it will happen soon or later (Clemen 1996, Keeney 1992, Russo and Shoemaker 1989, von Winterfeldt and Edwards 1986). Thus, an important step in any decision analysis for the objective NMFS stated in the 1995 Biological opinion would be to quantify what "high" and "moderate to high" likelihoods mean.

For example, the 1995 FCRPS Biological Opinion adopted a set of probabilistic metrics[3] to determine whether proposed FCRPS operations would avoid jeopardy. In the decision analysis conducted by PATH, the quantification was summarized as follows (Peters et al. 2000):

> *To meet this [no jeopardy] standard, an action must result in a "high percentage" of available populations having a "high likelihood" of being above the survival threshold level and a "moderate likelihood" of being above the recovery level. "High" and "moderate" likelihoods have been informally defined as being 0.7 for survival standards, and 0.5 for recovery standards. NMFS has defined "high percentage" of stocks as 80% of the available populations. For the cases in which we are focused on the seven Snake River index stocks, this means that for an action to be considered to have met the overall jeopardy standard, the action must result in six stocks having a probability of 0.7 or greater of being above the survival threshold and a probability of 0.5 or greater of being above the recovery threshold (p. 5).*

In the new draft Biological Opinion, NMFS continues to espouse the same "high likelihood of survival and a moderate to high likelihood of recovery" descriptor for the jeopardy standards (NMFS 2000, p. 1-10), but rejects the earlier metric developed to represent a "high" probability of survival. The new survival metric is (NMFS 2000) defined as follows:

A standardized metric of 5% probability of absolute extinction (no more than one fish returning over the number of years in a generation) within 100 years…is reviewed in assessing whether the species has a high likelihood of survival under the proposed action...a 24-year period is also reviewed because the range of uncertainty around the 100-year metric is quite large and because there is potential to further modify the action in the near term through the adaptive management process (pp. 1-10 to1-11).

The recovery standard is defined as a "50% or greater likelihood of meeting the recovery abundance level in the specified time period" (NMFS 2000, p. C-4). It appears from CRI spreadsheets that the way this recovery standard is evaluated is by using a deterministic λ calculated via the ratio of the recovery standard population divided by the current population. The new λ appears to then be used to project the current population forward 48 years, assuming an exponential population curve. How, exactly, this is actually done, as well as how such a simplistic approach might be justified, is not clear. Text in the draft Biological Opinion appears

[3] In addition, then, as now, NMFS made clear that the assessment was both quantitative *and* qualitative. Decision analysis offers some useful methods for incorporating qualitative judgments into an assessment, some of which were used by PATH (e.g., the professional judgment elicitation effort called the "Weight of Evidence" process) and none of which have been used in the current NMFS effort. Simple logic dictates that either the quantitative and qualitative assessments must be consistent with each other, or discrepancies have to be resolved.

to indicate that Leslie matrices are used, but the analytical tools that have been made available by NMFS (version 9/7/00) do not use Leslie matrices in this way.

Aside from the lack of consistency between the written draft Biological Opinion and the analytical tools that NMFS has made available, the definitions of these metrics and standards in the draft Biological Opinion present a fundamental problem for at least two reasons: (1) the probability of absolute extinction is not clearly related to the survival objective. That is, the logic is missing that would show that a 95% probability of avoiding absolute extinction is adequately representative of the actual biological survival of the species; and, (2) the analyses of whether future actions will achieve the new quantitative survival standard (a 95% probability of avoiding absolute extinction over 100 years), let alone the recovery standard, are not adequately representative of the probabilistic metrics. That is, the analysis of future actions provided in the draft Biological Opinion does not adequately estimate the probability that these ESUs will avoid absolute extinction and achieve recovery in a statistically and logically valid way. Each of these issues is addressed in turn below.

V.1.1. The survival metric does not adequately represent the survival objective.

The metric used to evaluate the survival standard is absolute extinction as predicted by the modified Dennis model (NMFS 2000, section 1.3.1.2). The survival standard is that there must be no more than a 5% probability of 1 or fewer fish returning during any 5-year period sometime in the next 24 or 100 years[4]. This statement of the metric would appear to indicate that if the modified Dennis model predicts there is even a 1% probability that at least 2 spring and summer chinook future spawners will be alive somewhere (as immature adults in the ocean, for example), for at least one out of any five consecutive years, then the probability of survival is "high," and a course of action predicted to produce such a result would be an acceptable means of achieving the survival standard.

It is not stated whether this metric applies to the ESU as a whole, to each population, or perhaps to some representative portion such as the SRSSC index populations. The 1995 standard cited above required that 6 out of the 7 index populations meet or exceed the survival and recovery thresholds. Missing from the current metric is a definition of how many populations in an ESU would have to meet the standard in order for the ESU as a whole to meet it.

Although there is widespread agreement in the conservation field that a reasonable threshold for population viability is more than 1, there is little agreement on what that threshold might be for a particular population, time, or place. There is, however, widespread agreement that, because various kinds of stochasticity[5] add to the risk of extinction, a meaningful population survival threshold — one that will actually ensure that a population survives — must include some

[4] Other metrics are also discussed and analyzed in the CRI, but the metric of absolute extinction in 24 or 100 years is the one defined as *the* standard in the draft Biological Opinion.

[5] The four main sources of stochasticity are: environmental, genetic, demographic, and catastrophic (Burgman et al. 1993, Thompson 1991).

margin (ISAB 1999).

NMFS scientists have argued in the past that because of these stochasticities, as well as depensatory effects, survival thresholds for viable salmon populations should be set high enough to adequately account for such risks (Thompson 1991, Wainwright and Waples 1998). To address this concern, Botsford and Brittnacher (1998) used a quasi-extinction threshold of 100 chinook spawners, arguing that, for Pacific salmon, depensation occurs at around 100 females. One of the lowest thresholds anyone has used for Pacific salmon was Mundy's (1999) threshold of 15 spawners in a brood year, although, following established practice, he focused as well on rate of decline. Nicholson and Lawson (1998) defined quasi-extinction as ≤50 spawning coho in a basin. The smallest long-term (100 years or more) viability threshold recommended for Pacific salmon or similar species in a recent NMFS white paper is 500 adults; most recommended thresholds cited in that paper are on the order of 1,000 to 10,000 adults (Thompson 1991). Thompson's paper was discussing thresholds for viability, not extinction, but presumably if a population is not viable, then it would be impossible to meet the survival jeopardy standard. This is an important issue that should be considered in setting modeling thresholds.

Since 1995, the average spawner population size for every one of the SRSSC index populations has declined to below the threshold set by the Biological Requirements Work Group (BRWG) (150-300 spawners) as part of the 1995 Biological Opinion process, and the CRI has shown that the rate of decline has actually been accelerating.

The implications of choosing a particular threshold are significant, and the logic of the choice of threshold should be carefully articulated and be consistent with the available scientific evidence. The choice of an appropriate extinction threshold against which to assess survival probabilities is critical because different thresholds will carry different implications about the risks of delaying actions that may favorably affect salmon survival, as well as the overall conclusion that the RPA in the draft Biological Opinion, along with offsite mitigation efforts, is indeed a reasonable means of satisfying the jeopardy objective.

The argument that NMFS offers in support of the absolute extinction threshold it has adopted for assessing survival in the draft Biological Opinion is that data are often not available for some of the ESUs, and that absolute extinction is thus the only threshold that can be standardized across all of the salmon ESUs in the Columbia River basin (CRI 2000b). The logic of this argument cannot be sustained as part of a rational decision regarding Snake River chinook salmon populations because there are considerable data available for these index populations, along with population thresholds defined by the BRWG and used in both the 1995 Biological Opinion and the PATH analysis. The availability of this information makes the decision not to use a more reasonable biological survival threshold in order to achieve standardization unsupportable.

NMFS also cites concerns expressed by the Independent Scientific Advisory Board in a review of the A-Fish Appendix, and by a panel that reviewed the BRWG work, but those concerns are mis-characterized in the draft Biological Opinion. Neither the ISAB nor the BRWG reviews were suggesting absolute extinction would be a more appropriate threshold; in fact, in a review conducted specifically of the CRI, the ISAB (1999) argued that the CRI's standard of absolute extinction for only one brood line was too low. What the ISAB recommended was that NMFS "consider a range of definitions for the quasi-extinction level, present the rationale for each, and calculate the probability of quasi-extinction associated with each" (ISAB 1999, p. 15). Since then, NMFS has instead moved the standard even lower, namely to absolute extinction for an entire population, not merely one brood line.

A second, similar problem with the choice of the survival metric in the draft Biological Opinion is that the CRI actually advocates evaluating the probability of population *decline* to assess the likelihood of survival, rather than (or perhaps in addition to) absolute extinction, not least because it would be more consistent with prevailing scientific practice in conservation biology. Thus the CRI states (CRI 2000a):

> "**In general, if "extinction peril" is a quantity to be estimated, because of the poor quality of existing data the CRI feels that the probability of a 90% decline is the best measure of risk**" (p. 15, emphasis in original; see also similar assertions on p. 16 and p. 49).

Yet the draft Biological Opinion instead relies only on the less-conservative absolute extinction threshold in assessing the survival standard. There is a significant difference in the conclusion that would be reached using probability of decline, rather than an absolute extinction threshold, as a survival metric. For example, for the case where the potential impacts of hatchery fish are assumed negligible[6], the modified Dennis model estimates that the risk of absolute extinction within 24 years is nonexistent[7] for all the populations except Sulphur Creek, whereas the average risk of a 90% population decline is close to 10% (Table 1).

It is worth pausing to think for a moment about what "90% decline" means for populations that have dropped to 0 twice in the past 5 years, such as Marsh and Sulphur Creeks, and whether concluding that there is no chance of extinction for one of them makes sense.

Table 1. Probability of absolute extinction or 90% decline within 24 years. Source: CRI 2000, Tables B-4 through B6, 5-Sept. 2000 draft; CRI spreadsheets 7-Sept. 2000 draft. Geometric means were calculated for Marsh and Sulphur Creeks by replacing spawner counts of 0, with 1, because the geometric mean is undefined for 0-values.

Snake River index population	Current 5 year geomean spawners (1995-99)	Probability of absolute extinction	Probability of 90% decline
BEAR VALLEY	88	0.00	0.07
IMNAHA	214	0.00	0.05

[6] The modified Dennis model extinction risk estimates are even higher when interactions with hatchery fish are increased in the model (CRI 2000b).

[7] Probabilities were shown as fractions to two decimal places in the CRI, and it seems reasonable to assume that NMFS did not mean to imply that a probability of 0.00 means there is absolutely no probability of extinction

JOHNSON	67	0.00	0.01
MARSH	13	0.00	0.13
MINAM	111	0.00	0.06
POVERTY	189	0.00	0.05
SULPHUR	14	0.05	0.15

Since the CRI (2000b) argues that the probability of decline is not only less vulnerable to error but also a more easily justified metric for likely survival than is absolute extinction, it seems inconsistent for the draft Biological Opinion not to use it, particularly when a population decline metric would provide such a sharply different picture of the likelihood of survival for the species.

To be consistent with the state of the science in conservation biology, the choice of standard should not be between two simplistic measures. The CRI's analysis of two metrics at least provides a broader (though still inadequate) context in this sense than does the draft Biological Opinion. Neither the World Conservation Union (IUCN) nor NRC would advocate using only the rate of decline metric exclusively (let alone the absolute extinction metric alone); instead, the current IUCN approach is to consider several characteristics of a population, particularly its current state (e.g., population size over some period of time), rate of decline, distribution (number of sub-populations, connectivity, area, etc.), and any acceleration evident in that rate of decline (Musick 1999).

The draft Biological Opinion's choice of standard is particularly difficult to understand because NMFS has been part of a group working to define extinction risk metrics, which includes the American Fisheries Society, the IUCN, the Canadian Department of Fisheries and Oceans, the Committee on the Status of Endangered Wildlife in Canada, and the Japanese government. This group is also working to define criteria appropriate for defining the risk of extinction for marine fishes. At this time, the American Fisheries Society criteria under consideration are: rarity, specialization in habitat requirements, endemicity or restriction to a small range, and population decline. In a recent paper describing the work of this group, there was no mention of any consideration of any metric even approaching "absolute extinction" (Musick 1999). The draft Biological Opinion does not address its departure from these approaches to evaluating the larger picture in the draft Biological Opinion, even though these approaches reflect the state of the science not only internationally, but in other parts of NMFS as well.

V.1.2. The analyses do not adequately represent the metrics.

The standard tools for analyzing extinction risks and population dynamics for salmon populations have increasingly been some stochastic variations on the Ricker or Beverton-Holt spawner-recruit themes (e.g., Hilborn and Walters 1992, Ludwig and Walters 1985, Marmorek et al. 1996), or Monte-Carlo Markov or Leslie matrix models (e.g., Botsford 1998, Higgins 1999, Ratner 1997, Nickelson 1998). Part of the reason that the Dennis model has had less to offer for salmon is that the Ricker model and its variants are familiar to fisheries modelers, and can be just as simple as the Dennis model — simpler, in fact, than the modified Dennis model being used by NMFS now — but can nonetheless handle density dependence, harvest rules, carrying capacity, and other characteristics of salmon population dynamics, whereas the Dennis model cannot. The Ricker model and related variants do not have a problem with age structuring, which has turned out to be problematic for the CRI's Dennis model. Markov or Leslie matrix cohort models are flexible enough to handle any of these complications that the Dennis model cannot, but they do have the problem of requiring stage-specific demographic and environmental parameter inputs. As is the case for the Dennis model, spawner-recruit models do not necessarily require stage-specific data. There are widely available and well-developed methods for making any of these tools stochastic.

These standard tools for analyzing fish populations are not used in NMFS' analysis. NMFS' analysis of whether the RPA will meet the survival and recovery standards, though evidently extensive and detailed, is convoluted and inconsistent, and ultimately not relevant to the metrics in some critical ways. The survival metric is a *5% probability of extinction within 24 or 100 years*. The recovery metric is a *50% probability of recovery within 48 or 100 years.* One might expect that the analyses would evaluate the decision alternatives directly in terms of how proposed actions or alternatives perform against these probabilistic metrics.

As it turns out, though, NMFS does not actually calculate these metrics for any alternatives other than a continuation of the status quo. NMFS does not have an analytic tool which can be used to make probabilistic estimates about different management scenarios, only some simple spreadsheet models and Leslie matrix models that can make deterministic predictions about long-term equilibria based on different scenarios, and another model — the modified Dennis model — that can make probabilistic extrapolations from the past. These models are based on different assumptions, though they share some structural fundamentals, and produce distinctly different results when they should produce equivalent results; yet NMFS treats them as interchangeable.

In effect, NMFS analyzes the RPA by assuming that the quasi-stochastic modified Dennis model is equivalent to the deterministic Leslie matrix model or other deterministic exponential growth equations, and that a future equilibrium average population growth rate calculated with the deterministic Leslie matrix, or with a simple formula in a spreadsheet, is the same parameter as a particular percentile of a probability distribution from the quasi-stochastic Dennis model. In effect, this approach is equivalent to borrowing a probability distribution from a quasi-stochastic regression, in order to invent a confidence interval for a deterministic life cycle model, or even a deterministic exponential growth model, of some different "what if" scenario. The population growth rates calculated with these different approaches are about 10% to 13% different from one another, a discrepancy first identified by the ISAB (1999) but still unacknowledged by NMFS. This discrepancy is important: a 10% difference in annual population growth rate will cause a 2.6-fold difference in population size after a decade.

Of particular concern is the fact that the extinction model produces significantly higher growth rates than the empirical data or the Leslie matrix models do. Population growth rate estimates produced by the Dennis model correspond to life cycle survival rates that are 22% to 29% higher than estimates based on empirical spawner-recruit data. For example, since 1980 the geometric mean recruit/spawner rate has been 0.74 for the 7 SRSSC index populations, a rate well below replacement (CRI spreadsheets 7-Sept. 2000 updates). Because the Leslie matrix models are calibrated to the recruit/spawner data, they also are based on these empirical estimates and thus produce equivalent average growth rates.

Yet the Dennis model population growth rates currently being used by NMFS in the draft Biological Opinion are equivalent to assuming that the geometric mean recruit/spawner rate has been 0.96 to 0.99 during that period[8], a rate near replacement and simply not credible in light of the available evidence. This is an optimistic platform indeed on which to base estimates of extinction risk and design recovery plans.

NMFS treats these different models as interchangeable, even though they produce such different results from one another, in order to, in effect, "borrow" the confidence intervals from the retrospective Dennis model, and apply them to the deterministic spreadsheet projections. NMFS took this unusual approach in order to be able to report probability estimates for the potential success of the RPA, when no probabilistic tools had been developed for doing so. The problem with this approach is that while it might be possible to conclude that there is a 5% probability of extinction associated with a stochastic model's median growth rate of, say, 0.9, based on historical data, that is not the same as concluding that a mean population growth rate of 0.9, calculated from a deterministic model of an alternate management scenario, means there is a 5% probability of extinction, particularly when the two models are not even calibrated and are using different measures of central tendency (medians and means) as well.

No calibration is carried out between the various deterministic and stochastic approaches, and yet the CRI carries the most optimistic results from the Dennis model over into the deterministic analyses without acknowledging that the step gains a 22% to 29% boost in life cycle survival due entirely to the discrepancy between the models, a discrepancy which would not be supported by calibration or validation against the available data.

There are two more detailed problems with the equivalence that NMFS assumes between the two approaches.

- **The population growth rates called "λ" from the various models are not the same statistical parameters**

The statistical parameters used to represent population growth in the modified Dennis model, the Leslie matrix, or the deterministic exponential growth models, are not equivalent, as NMFS evidently assumes. The population growth rate that NMFS derives from the modified Dennis model is a geometric mean from a retrospective quasi-stochastic model's probability distribution. The CRI renames this parameter λ and calls it a median, though Dennis et al. called it α and defined it as a geometric mean[9] (Dennis et al. 1991). Dennis pointed out that α is not the same

---

[8] The CRI's equation (CRI 2000b, p. 17) for converting λ's to R/S was used to calculate R/S ratios equivalent to the λ's reported in CRI's Appendix B, 5-Sept. 2000 versions.

[9] NMFS refers to this parameter as a median, but Dennis et al. describe it as the "finite or discrete growth rate of the geometric mean population size" of a diffusion process (1991, p. 126).

as the dominant eigenvalue, λ, from a stochastic Leslie matrix projection. Dennis et al. developed a different parameter which they called λ because it, unlike α, is supposed to be equivalent to the λ that would be produced by an equivalent stochastic Leslie matrix.

The CRI is using a deterministic Leslie matrix, not a stochastic one, and so there are actually three distinctly different parameters which the CRI treats as interchangeable when they are not: the geometric mean population growth rate from a quasi-stochastic Dennis model (α), the long-term average population growth rate from a stochastic Leslie matrix, equivalent to the average population growth rate from a quasi-stochastic Dennis model (λ), and a different long-term average population growth rate from a deterministic Leslie matrix (a different λ). The simple multiplicative or exponential growth calculations that the CRI is currently using to calculate life cycle survivals instead of the Leslie matrix models also uses a population growth rate equivalent to a deterministic long-term average (and not a stochastic median). The CRI thus renamed a parameter from one model so it appears to be the same parameter in a different model; but the parameters are not the same, identical label or not.

α from the Dennis model could, in theory, be used to extrapolate a population's trajectory into the future, assuming demographic variables will be unchanged, that demographic or genetic variability have negligible effect, that there is no density dependence (i.e., carrying capacity limitation at high numbers or depensation at low numbers), and that environmental variability remains the same. The variance estimate derived from the modified Dennis model's regressions can be used to produce probability distributions around the forecast, which can be used to estimate the probability of a population being some size, at some future point in time, assuming the trajectory continues smoothly on the same course.

However, in contrast, the long-term equilibrium growth rate λ's calculated by the CRI's deterministic Leslie matrices, or assumed in the deterministic exponential growth models, are different from the Dennis model's α in at least two important ways: they are calculated from *deterministic*, not stochastic analyses, and, as noted above, results that should be calibrated to each other are distinctly different.

In short, the CRI uses Leslie matrix λ's and some equivalents in simple exponential growth models to extrapolate deterministically into the future, but the only CRI tool that can provide probability distributions and thus extinction probability estimates[10] or recovery probability estimates is the retrospective Dennis model. Because it incorporates demographic parameters into the analysis, only the matrix model or simple spreadsheet exponential growth functions could be used to model scenarios different from an extension of the status quo, but the CRI's Leslie matrices are not stochastic and thus have no way to produce probability estimates.

There is no legitimate way to use the CRI's Dennis models to simulate population trajectories if the time period modeled includes changes in demographic, environmental, or genetic factors because doing so would violate the requirements of the Dennis model. The deterministic exponential growth equations used by CRI to project populations into the future for the purposes

[10] Leslie matrix models are flexible and as will be discussed later, could easily be stochastic as well.

of evaluating prospects for recovery are not only completely deterministic, but they share with the Dennis model a lack of any carrying capacity limitations.

- **Neither population growth rate (λ or α) is an adequate representation of population change in the near future**

The long-term equilibrium average growth rates calculated using deterministic Leslie matrix or spreadsheet survival and recovery models are not likely to be representative of what these populations will do in the next couple of decades because these decades will be a period of non-stationary change, either because the accelerating decline will continue, or, in theory, because of implementation of the RPA, if nothing else. The dominant eigenvalue from a deterministic Leslie matrix — λ_1 in textbook terms — represents long-term average population growth under stationary conditions. Short-term population growth under conditions of change is not generally at equilibrium, but instead tends to oscillate or fluctuate more widely in patterns (if there even are patterns) defined more representatively by the sub-dominant eigenvalues (λ_2, λ_3, etc.) rather than the dominant eigenvalue λ_1 (Burgman et al. 1993, Fox and Gurevitch 2000 in press). This distinction is important because, for example, the SRSSC index populations are already quite small, and short-term variability is a well-recognized key factor in extinction processes.

It might be tempting to assume, as the CRI does, that the long term average parameter at least can be used as a rough guide to how the populations would perform if more realistic population characteristics such as accelerations, thresholds, and stochasticities were taken into account, but this assumption is mathematically untenable. There is in fact no predictable relationship between any system performance under accelerations and stochasticities, and system performance if accelerations and stochasticities are ignored, particularly if thresholds and cumulative effects are involved. This is a fundamental principle of nonlinear mathematics that applies to predator-prey systems, physical systems, or any kind of system which has temporal, spatial, demographic, or other interacting mechanisms (Byers and Hansell 1992, DeAngelis and Post 1991, Holling 1976). For example, the long-term yield of wood from a dry interior forest will be greatly over-estimated if one ignores the risk of forest fires, driven by gradually accumulating fuel loads that eventually exceed a critical threshold, not to mention the stochastic nature of lightning strikes. No sane manager of such a forest would fail to consider such thresholds and stochastic events.

The population growth rate derived from a modified Dennis model's analysis of historical data is not representative of what these populations will do in the next couple of decades either, unless the status quo actually has been stationary since 1980 *and the population will continue to be at the same equilibrium into the future*. These assumptions could, in theory, be satisfied for an analysis that projects a continuation of a stationary status quo, but not for an analysis of the RPA. The RPA is, after all, a proposal to change the status quo and thus violate the stationarity assumption.

The near future matters for these populations. Two of the 7 SRSSC index populations have seen spawner counts of 0, twice each since 1994. CRI analyses, optimistic as they are, nonetheless indicate significant risks of extinction even over the next few decades, particularly if hatchery fish interact significantly with wild fish, or if the 90% decline metric is used.

To illustrate the trajectories that various NMFS models and the RPA assumes are occurring and will occur under the RPA+mitigation proposal, Figure 2 shows (1) the most recent empirical estimates of the population of the aggregate SRSSC population[11], (2) the trajectory predicted by the current Dennis model analyses, (3) the trajectories that would be produced with the target increases calculated from the Dennis model's assumed growth rates, starting in 2001, and (4) the trajectory extrapolated using Leslie matrix population growth rates calibrated to spawner-recruit data.

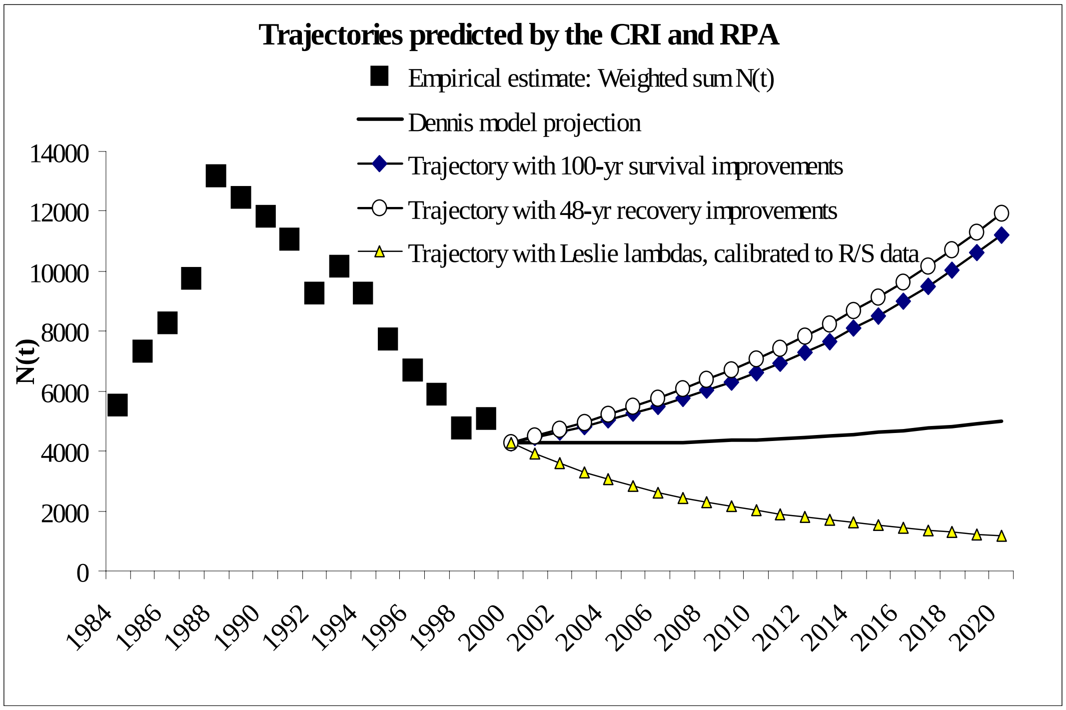

Figure 2. Empirical estimates and trajectories predicted by λ's calculated from empricial data, by the Leslie matrix model, and by the Dennis model, using 1980-1999 data; and by the improvement required to avoid extinction or 48-year recovery goal, assuming improvements are fully effective in 2001. λ's calculated to represent required improvements were based on Table B-8 (CRI 2000b, 9/5/00 version) and CRI spreadsheets dated 9/7/00.

The first question one might ask about Figure 2 is which of the trajectories seems the most reasonable extrapolation of the current trend. The most reasonable trajectory, not surprisingly, is obviously the one that is based on empirical spawner-recruit data ("Trajectory with Leslie λ's using R/S data"); but the extinction probabilities, and the target improvements required to achieve survival and recovery goals, are calculated via the modified Dennis model only. NMFS has not acknowledged this discrepancy nor justified such heavy reliance on a model that produces such questionable projections.

---

[11] The vertical axis uses the total population estimates *N(t)* rather than simply spawner counts; see CRI 2000b for descriptions of Dennis model modifications.

Secondly, one might ask if the extrapolation currently favored by NMFS seems reasonable. NMFS' favored trajectory is labeled "Dennis model projection." This projection (which uses population growth rates from the draft Biological Opinion update dated 9/5/00) diverts the obvious downward trend into a virtually constant average population size with no evident downward trend — truly an odd behavior for what is supposed to be a continuation of the current trend. Recall that the Dennis model should produce approximately a continuation of the established trend, not an abrupt step up. This unexpected model performance appears to be due to modifications the CRI has recently made to the Dennis model in the 9/5/00 updates, and has not yet been acknowledged nor explained by NMFS. This kind of pattern is not the way this model should behave, and before the model could be treated as credible, such a departure would have to be justified. Because NMFS assumes this abrupt step up does indeed represent the current trajectory, only a moderate upswing is projected to be required from that in order to meet the survival standard. A slightly larger upswing is then found to be required to meet the recovery standard.

A third point illustrated by Figure 2 is the effect of assuming that the increased survival rates due to the RPA and offsite mitigation will occur[12] in full in 2001, that these populations and their environments will immediately move to a new stationary equilibrium, that demographic rates and environmental variables are not only constant but will not vary with the new actions, and that the predicted result will persist, without variation, forever: an instantaneous improvement in the status quo put into full effect in 2001. At a minimum, this approach appears to assume that there are no interactions with other species (including humans), no environmental fluctuations due to la niña, el niño, floods, drought, Aleutian Low Pressure Index or North Pacific Decadal Oscillation, no high demand for electricity in the summer, and no delays or problems in implementation (even when delays are acknowledged in the text), among other things. The basis for assuming away this range of risks and uncertainty is not articulated. The unavoidable effect of these assumptions is that the picture presented of the present, not to mention the future, is highly optimistic and unjustified.

The summary message that Figure 2 illustrates is that NMFS' models and conclusions are not consistent with existing data nor an understanding of how political and natural systems realize or respond to change in the real world. The unrealistic population trends in Figure 2 produced by NMFS' model of the status quo, and of the improvements produced by the RPA, show how inadequate such an approach is for representing how a course of action might relate to the ultimate objective of avoiding jeopardy and ensuring recovery.

V.2. Performance standards as metrics

A common approach to making metrics more manageable is to define sub-metrics that are more directly representative of the details of the alternatives and objectives. Stratifying such sub-

[12] The discussion in the text of the Biological Opinion acknowledges and even sometimes defines implementation schedules that are not consistent with this fundamental modeling assumption; but it is the modeling results that are relied on the most heavily in justifying the RPA and offsite mitigation proposal.

metrics, as NMFS is proposing to do with the performance measures and standards[13], can be quite useful. As the All-H paper puts it (Federal Caucus 2000), performance standards:

> *…are the means for establishing the level of survival improvement in each stage of the salmon and steelhead lifecycle that are necessary for survival and recovery. Not only do performance standards create clear objectives, they provide flexibility to define the most efficient means of achieving the objectives. Performance standards have been divided into three tiers...Over time, compliance with these standards would be assessed through monitoring and evaluation. If progress toward meeting performance standards is insufficient, adjustments can be made, either in the actions implemented or in the allocation of survival improvements across the H's (pp. 46-470).*

NMFS uses the modified retrospective Dennis model to calculate the percent change in the geometric mean (NMFS calls it a median, as discussed previously) equilibrium population growth rate that would theoretically be required to meet the jeopardy standards for each population. For example, the target equilibrium growth rate increases required to meet the survival and 48-year recovery standards, calculated for the 7 SRSSC index stocks, average less than 6% (depending on what is assumed about hatchery impacts). These target "λ's" are transformed to a linear, additive series of life-stage survival rates in Excel spreadsheets, in order to calculate what percent increases in survival would be required at each life stage, to meet the overall life cycle target increases and thereby ensure survival and recovery (or stated negatively, avoid jeopardy).

Those increases of 6% or less in "λ" required to meet the survival and recovery thresholds translate into approximately 30% increases in life-cycle survival rates. Even then, these target increases are medians: by definition, the probability of the true value being less than a median is 50%. The survival standard requires a 95% probability that the survival rate will be greater than the target rate, and the best a deterministic model can do by way of probability estimates is to estimate a median.

These target increases in life stage survival are to be used to identify measures in each of the H's which are *assumed* to lead to the increases in life stage survival rates, which then become the performance standards for each life stage.

Many of the actions to satisfy these performance standards have not as yet been defined, and many of them are not under the control of the decision-makers because they are owned or managed by private, state, or other entities. The draft Biological Opinion states that all of the offsite mitigation efforts are under control of the Action Agencies ("Offsite enhancement includes only measures that are within the current authorities of the Action Agencies", NMFS 2000, p. 9-16), but the All-H paper states that, "The federal agencies cannot solve this problem alone, or by acting unilaterally. Strong action by state and tribal governments, local authorities, and other

[13] "A performance measure describes a population, life-history stage specific, or human activity-specific biological condition. A performance standard is a value of a performance measure that has been identified as a management goal." (Federal Caucus 2000, p. 46).

participants must occur for recovery to succeed. All parties must coordinate efforts to fully realize benefits to species in decline" (Federal Caucus Executive Summary 2000, p. 2), so it is not clear even who has the power to make the RPA operational. With so much of the RPA undefined or defined in so many different ways, it is difficult to find convincing the arguments that it will avoid jeopardy.

Just because an outcome is said to be "required" or a "standard" hardly means it can be counted on to satisfy an objective. A structured decision would maintain a clear distinction between actions proposed, and the metrics used to evaluate those actions. NMFS does define some actions to aid juvenile and adult passage migration, which NMFS acknowledges fall short of meeting the jeopardy standards. But NMFS then assumes some undefined suite of actions will achieve the shortfall as measured by the "performance standards" metrics. The logic of this proposal is convoluted, but in essence it appears to presuppose the desired outcome, in defining the alternative. Then the analysis is used to show that the alternative will achieve the outcome, as presupposed. That "wishful thinking" trap is not uncommon when decisions are not carefully structured, but it is avoidable if decision structuring tools are used to help keep the analysis on a logical track (Russo and Shoemaker 1989, von Winterfeldt and Edwards 1986).

The section in the draft Biological Opinion titled "Assessment of standards related to status of ESUs" (p. 9-22) states that the checkpoints will consist of evaluating overall performance of the populations by evaluating whether λ is greater than 1.1, at the 2005 checkpoint[14], using the modified Dennis models. Choosing a standard of 1.1 seems somewhat arbitrary, since there is no relationship between this target 10% annual population growth rate, and the target population growth rate increases calculated with the Dennis model. Be that as it may, however, on page 9-19 of the draft Biological Opinion, it appears that the midpoint evaluations will be based on more general considerations — so it is hard to know what, exactly, these checkpoints will be. Presumably a variety of evaluations and considerations will be taken into account, but what will happen at the checkpoints remains ambiguous and inconsistent, at least as articulated in the draft Biological Opinion.

Aside from the inconsistencies and ambiguities about the checkpoints, the most concrete measure of population performance that NMFS is planning to use is evidently the overall Dennis model calculation of equilibrium median population growth rate, and probability of extinction.

There are several slips from the Dennis cup to the performance standard lip, even setting aside the problems with using the Dennis model for these populations discussed above.

First, as discussed above, changes in the Dennis model's estimate of the population growth rate necessary to reduce retrospective equilibrium median probabilities of extinction to 5% are not equivalent to future average, equilibrium population growth rates calculated with a Leslie matrix, let alone an additive spreadsheet table. The basis for assuming this equivalence is not set forth in the draft Biological Opinion or the available CRI analyses.

[14] The checkpoint in 2008 will use a 3-level criterion: $\lambda > 1.1$, $\lambda < 0.95$, and λ between 0.95 and 1.1, to determine further action.

Second, the performance standards are based on the most optimistic combination of assumptions possible. That is, all the analytical models relied on by NMFS assume that the actions in the various H's are certain to provide the most positive change in life-stage specific survival rates, and that they will happen immediately, without variation, with no negative impacts, for 24 or 48 or 100 years.

Third, none of the NMFS analyses allows for a transition period. Improvements in, say, egg-to-smolt survival, or estuary survival, happen instantly and without variability. Even hydrosystem improvements that were promised in 1995, and have yet to be implemented, are assumed already effective. These assumptions of immediate effects are inconsistent with the descriptions in the text, which acknowledge, for example, that the additional hydro improvements will not be fully implemented until 2010.

In short, because the analysis does not include these important characteristics of natural systems — climatological, environmental, and demographic variability, time lags, and the fact that constant, smooth exponential population growth does not occur for 100 years without some density dependence or carrying capacity limitations — even if they could somehow be put into full effect and fully achieved tomorrow as defined, the performance standards will surely be inadequate to achieve what the ESA says they must. These issues are brought up a number of times in the draft Biological Opinion, but not incorporated into the conclusions.

The above observations are not limited to biological considerations only. As the draft Biological Opinion acknowledges, in order to improve survivals for each life stage, first the federal, state, and private programs have to be planned, funded, and implemented; then those programs will have to actually improve hydro, habitat and hatchery attributes as expected; and finally, those changes will have to increase the life stage-specific survival rates for the ESA-listed populations. The matrix and spreadsheet analyses do not take these inescapable facts into account, and the Dennis model analyses cannot. These biological and pragmatic considerations apply to all 12 ESUs.

Finally, to be useful in an ongoing decision or risk management process, changes in metrics have to be detectable. To tell us at the 5-year and 8-year checkpoints whether the RPA is working, the proposed actions will have had to have improved the recruit-per-spawner survival rate enough for the CRI's modified Dennis model to reflect those changes. This fundamental calculation is itself impossible because of these populations' 5-year generation length and the way the spawner-recruit data are modified in the CRI's version of the Dennis model.

Because the modified Dennis model uses 5-year weighted sums to estimate each year's population, and two such population estimates are required to estimate the stationary population growth rate (more, in fact, are required to estimate the variance), it could not be updated with adequate validity until at least 6 years had elapsed since significant changes were made, and the system had been given enough time to settle into stationary equilibrium. Averaging in spawner counts from before implementation of the RPA would simply contradict not only the assumptions of the model, but also the assumption to be tested — namely, that the RPA has made significant changes to life cycle survival. The approximation being used by the CRI for manipulating the spawner-recruit data does not use spawner-to-survival survival rates to calculate λ[15], but it would still take a minimum of 6 years to develop even a single additional data point that could be used to update the analysis to reflect the changes. The reason for the delay is that the Dennis model assumes stationarity, which means previous data points from the 1980-1999 period would not be useful because the whole point of the RPA is to implement immediate, significant change, thus violating stationarity. If there really is no change, then previous data points could be used, but if there is no change, it would already be clear that the RPA had failed anyway.

Thus, by 2005 there will not even be one data point for the Dennis model to use to recalculate the new λ. If the more rigorously-defined method for estimating population size in the modified Dennis model is used (which does include annual spawner-to-spawner survival rates), instead of the approximation favored by the CRI (which does not include annual spawner-to-spawner survival rates) (CRI 2000b) then it would take a minimum of 11 years for one additional data point to be added to the Dennis model. The Dennis model calculations of λ would clearly be useless for assessing life-cycle survival improvements against the "$\lambda > 1.1$" standard offered in the draft Biological Opinion.

If the question is whether, under NMFS' preferred option, the Snake River spring and summer chinook ESU has a high probability of surviving, or moderate to high probability of recovery (i.e., the jeopardy standards) — and that is the question NMFS set out to answer — then NMFS' analyses do not, and cannot, adequately address it in the time frames NMFS has set for the 5-year, 8-year, and 10-year checkpoints. The approach of waiting another 5 to 10 years, while the RPA is put into effect, and relying on an analytical "safety net" to detect a mistake, simply is not supported by logic. In short, there is no safety net.

VI. Alternatives

Before any strategy can be usefully analyzed, it has to be defined clearly enough that decision-makers understand what, exactly, they are analyzing. If the alternatives are not clearly defined, then the analyses might not be applied to the actions that would actually be undertaken. One poorly-defined alternative might look better than another in an unstructured context, but how that preferred alternative would perform in the real world would not have been properly evaluated. Only after the decision was put into effect, and did not turn out as hoped, would it become clear that the decision-makers did not make full use of available information when they defined and analyzed the alternatives. Decision structuring helps decision-makers avoid such unpleasant surprises because it requires them to carefully define what, exactly, it might mean to carry out each alternative action that is under consideration.

In particular, it makes little sense to evaluate alternatives that are not likely to be feasible, or alternatives for which feasibility has not been considered. Every analysis of FCRPS management options done since NMFS issued its 1995 Biological Opinion has used a different

[15] Spawner-to-spawner survival cancels out of the CRI's equation 4.4B (CRI 2000b p. 27), which is an approximation to the actual definition provided by Holmes.

set of alternatives. Until the CRI was launched, at least those alternatives had been clearly defined, and their feasibility had been assessed to some extent. PATH, for example defined 7[16] feasible alternatives in some detail, so that they could be analyzed using standard decision analysis methods. STUFA applied the CRI matrix model to 6 alternatives, which were based on those analyzed by PATH (Oosterhout et al. 2000). The 7 alternatives defined and analyzed by PATH and to some extent by STUFA are the only ones defined and developed to date with feasibility in mind. The alternatives evaluated in the draft Biological Opinion have not even been substantively defined yet, and NMFS at least acknowledges that feasibility has not yet been taken into account either.

So the obvious question is this: What are the actual actions analyzed in the draft Biological Opinion, and how do they relate to what is required by the ESA? The CRI did not define or analyze any management alternatives. Instead it analyzed hypothetical equilibrium increases in survival rates, which would presumably be achieved by some unidentified actions, and the feasibility of which, in the timeframe required, has not been determined or incorporated into the decision. The alternative actions actually described in the draft Biological Opinion consist only of measures to increase juvenile passage survival and adult passage survival. The All-H paper describes the beginning of a wide variety of planning and prioritization efforts, including an emphasis on developing performance standards by life stage — but no clear alternative actions whose potential outcomes are, or can be, assessed. It may be useful to study potential results of hypothetical increases in egg-to-smolt and estuary survivals as was done in the CRI analyses, if the models have been validated to a significant extent (which the CRI analyses have not), but hypothetical, infeasible, survival increases do not constitute alternative actions whose probability of meeting the jeopardy standards can be assessed with adequate confidence or scientific rigor.

In a rational decision process, analyses would apply directly to the alternative actions described in a biological opinion. But because the alternatives are not defined in this draft Biological Opinion, the ability of the analyses to demonstrate that the proposed actions meet the objectives is missing.

VII. Variability, uncertainties and error

Common sense indicates that it would be irrational to make an important decision in the face of considerable uncertainty without taking into account what can go wrong. This common sense insight has led to development of an extensive body of science about how to manage decision making under conditions of uncertainty (Clemen 1996, Hilborn and Walters 1992, Holling 1978, Howard 1968, Howard 1989, Keeney 1992, Keeney and Raiffa 1976, Punt and Hilborn 1997, Russo and Shoemaker 1989, von Winterfeldt and Edwards 1986, Watson and Buede 1987). This literature shows that to some extent, one can reduce the effects of uncertainty by such tactics as implementing a diverse portfolio of measures, or planning for the collection of information during implementation and creating options that can be exercised on the basis of that information. Uncertainty that cannot be managed in this way can be more formally incorporated

[16] Or 6, depending on how they are counted (Peters and Marmorek, in review, Canadian Journal of Fisheries and Aquatic Sciences, 2000).

into a decision by encoding uncertainties, variability, and error in probability distributions and sensitivity analyses.

A great deal of research has consistently shown that most decision-makers are prone to *over*estimate the appropriate level of confidence in their own knowledge and analytical abilities, and that they are prone to *under*estimate the unpredictability of the real world (Keeney 1992, Russo and Shoemaker 1989, von Winterfeldt and Edwards 1986). These decision-making failures arise from problems of underestimating or mischaracterizing the ranges and effects of *variability*, *uncertainty*, and *error.*

Variability is an inherent characteristic of a system, which cannot be reduced through study or measurement but may be reduced (or increased) through human intervention. It is sometimes called "aleatory uncertainty" or "stochastic variability." For example, environmental variability includes ocean cycles, droughts, floods, and heatwaves. Environmental variability is assumed to be the source of the regression error in the Dennis model. The range of this variability is used to estimate probabilities of extinction if past conditions continue into the future. The Leslie matrix and spreadsheet analyses, however, are deterministic and set environmental variability to zero.

Uncertainty is a level of ignorance, or lack of knowledge, about a system's parameters, functional relationships, and patterns of variability. It may be reduced by study or research. Uncertainty is sometimes called "epistemic uncertainty" or "degree of belief" (Vose 2000). An example of the kind of analysis of uncertainty that can be useful is the CRI's analysis of how impacts from hatchery fish may affect extinction probabilities for the three index populations of SRSSC that may include hatchery fish. No one knows how much impact hatchery fish have had on wild populations in the past, nor how much impact they might have in the future. So the CRI calculated extinction probabilities assuming four different levels of historical impact from hatchery fish, in order to illustrate the potential impacts on the extinction calculations and required survival increases.

Error is due to measuring the wrong thing, measuring or estimating it wrongly, not measuring it at all, or drawing the wrong inferences. Where a model is employed to aid in decision-making, one of the most important sources of error is often model structure, which is why it is so important to understand how results might be different for different model structures (Hilborn and Mangel 1997). Error analysis is often carried out via sensitivity analysis of one kind or another. For example, CRI models assume stationarity, linearity, and lack of density dependence. The CRI did analyze to some extent whether assumptions of stationarity and lack of density dependence are justified. The conclusion reported was that stationarity is probably not supported for Snake River spring and summer chinook and that there is not enough evidence with such low populations to determine whether density dependence would be a factor. But the CRI uses models that rely on these assumptions anyway. As a result, the target population growth rates for these ESUs are based on models which the CRI's analyses indicate are probably inadequate, and they do not include any margin to account for such errors.

VII.1. Risk Assessment: Addressing Variability, Uncertainty and Error

The science of risk assessment is not about deterministic, expected values such as means or medians, it is about what is *not* expected: "When we calculate the mean and confidence limits from a set of numbers, we are looking at the central tendency of our data. The behaviour of outliers is largely ignored. Estimating risk involves calculating the chances of extreme events" (Burgman et al. 1993, p. 59). Research has repeatedly shown that predictions based on historical averages tend to produce overly optimistic conclusions because they underestimate the impacts of uncertainty, ignore the potential for errors, and fail to take into explicit account the well-documented unexpectedness or variability of natural systems (Burgman et al. 1993, Glickman and Gough 1990, Hilborn and Walters 1992, Vose 2000).

This approach chosen by NMFS is a clear example of what Hilborn and Walters (1992) were referring to when they observed:

> *…we believe that one of the reasons that major fisheries management disasters occur with discouraging regularity is the traditional inability and unwillingness of fisheries management agencies to deal directly and explicitly with uncertainty (Preface, p. x).*

There are 4 ways in which NMFS has failed "to deal directly and explicitly with uncertainty."

1. NMFS uses models that assume demographic, genetic, and observational variability and errors are insignificant.

One of the assumptions inherent to the modified Dennis model on which NMFS and the CRI rely in the draft Biological Opinion is that demographic, genetic, or observational variability and errors can be ignored. CRI acknowledges that this assumption is not satisfied as defined in Dennis' paper, and that (CRI 2000b):

> *…observation error in run-reconstruction data is likely to be extremely large, probably larger than 25%. This is especially evident when one realizes that often there are less than ten spawners (maybe even as few as one or two spawners) as the denominator of recruit per spawner ratios; a miscount of only one or two fish at such low population sizes can easily yield errors of 50 to 100% (p. 20).*

If only one or two spawners are indeed the entire spawner population counted, that would imply either that the population is so small that using such analyses to evaluate risk of extinction is superfluous; or the survey methodology is flawed. Either way, basing population predictions and management goals on data from such small populations, or flawed data, means that management decisions should incorporate a significant margin to account for the considerable uncertainty in the predictions. Miraculous things can sometimes be accomplished with statistics, but the CRI's argument that there is a statistical way around such questionable data, on such small populations, is not convincing, for at least two reasons: (1) the model results on which the argument appears to be based are not credible, when compared to available data (see Figure 2); and (2) although results have been made public for new analyses, the full explanation of, and justification for, the most recent modifications (9/7/00 Excel spreadsheets) have not been completed. Modifications which were advocated in the available version of the CRI report (dated July 17, 2000) produced significantly lower λ's, and higher probabilities of extinction, than are now being reported; and emails and conversations with CRI scientists[17] appear to indicate that the original Dennis method for calculating λ (or what Dennis calls α) has been adopted again, even though the July 17 CRI criticized the method apparently being used again now as being too sensitive to observation error. Thus, the potential impacts of demographic, genetic, or observation error have not been satisfactorily addressed in the CRI models on which the draft Biological Opinion is based.

2. NMFS uses models that ignore the best science available on risks to small and endangered populations.

Another problem is the application of exclusively exponential growth models of population projections. Exponential growth models are based on the simplest of assumptions about how a population might grow or decline, and population dynamics texts only use them as introductory concepts. They are rarely if ever used for modeling real populations (e.g., Burgman et al. 1993, Hilborn and Walters 1992). The most basic reason that scientists modeling fisheries use slightly more complicated models that can include carrying capacity effects is that, at the high end of the range, real populations cannot grow without limit forever, and at the low end of the range, real populations approaching extinction exhibit depensatory patterns that are not well modeled with exponential (nor probably any other currently recognized) functions. Over the near-term, it seems reasonable to assume that the index populations in good habitat are small enough that they are not in danger of exceeding freshwater carrying capacities anyway; but because the recovery analyses carry deterministic exponential population growth projections out 100 years, ignoring carrying capacity (not to mention uncertainties) is problematic.

On the low end of the exponential growth range, the fact that small populations become increasingly vulnerable to genetic and demographic variability is well recognized as a contributing factor in extinction processes (Allee et al. 1949, Burgman et al. 1993, Thompson 1991), and, as the CRI has noted, it is obvious that basing estimates of total population sizes 100 years in the future, on highly questionable data derived from spawner counts alone, is problematic. The CRI's extinction calculations, and target increases in population growth on which the performance standards are based, rely on a model that is structured to ignore such critical sources of variability and error. Multiple sources of variability and error mean that the variance used to calculate extinction risks in the CRI analyses is even more severely underestimated.

3. NMFS' models assume that increases in juvenile and adult passage, egg-to-smolt, and estuary survivals will happen immediately, in full, permanently, and without variability.

The most important failure associated with the step from analyzing population trends of historical data with the Dennis model, to projecting future trends under different management options with the Leslie matrix, may be the assumption that the hypothesized survival

[17] This point refers to an email from Dr. Holmes to Dr. Oosterhout dated Sept. 21, 2000; and explanations presented at CRITFC in Portland with Eli Holmes, Tom Cooney, Chris Jordan, from NWFSC. The explanations were documented in a report provided to us by Eli Holmes titled "Estimating risks for declining populations: salmonids as an example."

improvements happen in full and immediately, are constant, and are not subject to uncertainty or variability of any kind.

In a deterministic Leslie matrix model, or spreadsheet exponential growth model, like the CRI's, the variability of the future is assumed to be zero. Assuming that the variability in the future is zero means it is actually impossible to estimate the probability of anything, including extinction or recovery, because estimating the probability of extinction depends on the variance (Burgman et al. 1993, Thompson 1991).

What could NMFS and the CRI have done instead to address variability, uncertainty and error? They could have explicitly incorporated uncertainty about the future (including variability and the impacts of errors) into the analysis. NMFS says that the reason they chose not to conduct stochastic analyses on the future is because developing such models is time-consuming, and it is difficult to determine what the variability should be for matrix parameters (CRI 2000b). The obvious response to the latter assertion is that choosing to assume zero variability is no more defensible than assuming questionable variability.

Moreover, regarding the first assertion, one might wonder, "time-consuming compared to what?" The arduous manipulations and questionable linkages between the one stochastic model NMFS has — which only applies to an extension of the status quo and does a poor job of replicating that (see Figure 2) — and the prospective, but simplistic deterministic models NMFS is now using to analyze future scenarios such as the RPA, are highly problematic. The linkage between the retrospective stochastic model and the prospective, but simplistic deterministic models NMFS is now using to analyze future scenarios such as the RPA, are highly problematic. STUFA (Oosterhout et al. 2000) had already developed and to some extent validated stochastic models which were offered to NMFS; these models included the stochastic expressions for key variables, which were developed primarily from PATH sources. NMFS thus decided not to make use of the best data and analysis of variability and uncertainty that were available, even though the data and analysis had been extensively peer-reviewed, analyzed, and validated by, among others, state and federal agencies and tribal biologists with whom NMFS was supposed to be collaborating.

4. NMFS failed to carry out a credible risk assessment.

The field of risk assessment offers well-established ways to analyze and deal with unexpected variability and what can go wrong, and to build realistic assessments into decision analyses and management plans. Risk assessment is the science of assessing the probability that uncertain adverse events will occur, and the probability that impacts will be of particular kinds and at particular levels if they do (Vose 2000). Risk assessment requires more than fitting a regression through past data, as the modified Dennis model does. Regressions of one kind or another are sometimes used as forecasting tools, but decision analysis or risk assessment requires much more extensive consideration of mechanisms that could impact the outcomes.

For example, standard steps in a professional risk assessment are (Vose 2000):

1. Identify the risk to be analyzed.

2. Define the factors and mechanisms that could make the risk more or less likely to occur, or

make the impacts larger or smaller.

3. Analyze the uncertainties associated with the causative factors or mechanisms and their potential impacts.

4. Develop a risk management strategy (control, detection, mitigation).

5. Communicate with decision-makers and stakeholders.

Here is how the approach taken by the CRI and reflected in the draft Biological Opinion compare to this standard risk assessment methodology:

Risk assessment step 1. Define the risk to be analyzed.

The risk to be analyzed is the risk of not meeting the jeopardy standards. The metrics on which the analysis are based are, as discussed previously, the 5% probability of 24-year and 100-year absolute extinction metric, and the 50% probability of recovery in 48 years or 100 years metric.

Ordinarily, in evaluating serious risks, metrics are chosen that will be highly sensitive to risks: canaries in mineshafts, for example, because they are more sensitive to risk mechanisms than miners. A professional risk analyst would not generally chose an equilibrium average value such as λ as the focal point, because analyzing risks requires analyzing potential failure mechanisms and their potential impacts (Burgman et al. 1993, Glickman and Gough 1990, Vose 2000). By definition "equilibrium" means the system is relatively insensitive to perturbations — mathematical or in the real world. Equilibria and averages (or geometric means or medians) hide the variability of system responsiveness, and make risks more difficult to detect. Relying on models that assume stationary conditions means the analytical tools are inadequate to the task of analyzing the most important system characteristics from a risk perspective.

In short, NMFS states that an average (or geometric mean, or median) long-term population growth parameter called λ is an appropriate and widely used parameter (CRI 2000a and 2000b), but does not justify choosing a response variable that likely produces such optimistic and misleading results when both data and tools for a more realistic risk assessment are available, at least for a number of listed ESUs, particularly Snake River spring and summer chinook index populations.

Risk assessment step 2: Define the factors and mechanisms that could make meeting the jeopardy standards less or more likely to occur.

The risk mechanism for salmon survival that has received the most attention is the hydrosystem. One of the risk factors associated with this mechanism that others have identified is extra mortality or delayed mortality (Bouwes 1999, Marmorek 1998, PATH Scientific Review Panel 1998, Peters 2000). The CRI does include some discussion of extra, or delayed mortality, as well as some analysis of extra mortality related to the hydrosystem, though the analysis is superficial and ambiguous. However, other than identifying some potential implications of extra mortality, other hydrosystem risk factors are not incorporated into the analysis as they would be in a standard risk assessment. For example, analyses in the draft Biological Opinion assume that passage survival improvements *assumed* in the base case and promised in the RPA are certain and not subject to variability, uncertainty or error, even during a transition period. A standard risk assessment would devote considerable attention to what could go wrong with these assumptions, with particular emphasis on failure mechanisms and potential impacts, particularly those that can be recognized from past experience.

Another set of risk mechanisms of interest arise from the ocean, particularly as they affect survivals when juveniles first arrive in the estuary and near-shore environment. Ocean conditions also play a major role in the weather throughout the Columbia River Basin, including el niño and la niña events, droughts and floods. Because the matrix and spreadsheet analyses that the CRI uses assume constant, equilibrium, average growth rates in perpetuity, NMFS does not — and with the current tools it employs cannot — include such considerations in the analysis, though they do note in passing that, "The possibility that ocean conditions may improve relative to our current period, or deteriorate even further, may need to be considered when examining the policy options for certain ESUs." (CRI 2000b, p. 4).

Risk assessment step 3: Analyze the uncertainties associated with those risk factors and mechanisms.

NMFS has identified similar lists of critical uncertainties in the CRI and the draft Biological Opinion. For example, although it varies even within the CRI, in one place, the CRI (2000b) says this about critical uncertainties:

> *First, while better estimates are becoming available, the magnitude and mechanisms of indirect mortality associated with the hydrosystem or transportation have yet to be conclusively defined. Second, quantitative links between habitat and hatchery management actions and salmon productivity have not yet been established. Finally, the role of changes in ocean conditions in producing patterns of survival is not well understood; nor is the manner in which ocean conditions are likely to change in the future (p. 99).*

In the draft Biological Opinion, the critical uncertainties are identified as delayed mortality associated with the dams for juvenile as well as adult fish, feasibility of implementing freshwater habitat measures, magnitude of biological response to habitat actions, and effectiveness of hatchery supplementation.

The discussion in the CRI and the draft Biological Opinion about the importance of such uncertainties is substantive and important. Moreover, NMFS acknowledges the need to take action even in the absence of certainty about these topics, and says it is developing analyses and plans to help resolve them. More research is not going to resolve these uncertainties; experimental management might but only after 10-20 or more years of implementation and robust experimental designs (Peters et al. 2000). The draft Biological Opinion does acknowledge the potential for impacts from most of these uncertainties, and the CRI even analyzes some of them; but the ultimate conclusions and the analyses on which they appear to be based do not, even though there are well-established mechanisms to incorporate these uncertainties and potential errors into both the analyses and the conclusions.

For example, if history is any guide, the most reasonable assumption regarding implementation of federal, state, and Congressional actions listed under the "shared strategy" section of the draft Biological Opinion is that a substantial portion of these actions will not occur, or will be delayed, or will be only partially implemented. Even a logical analysis of the potential consequences of such failures could provide a clearer picture of likely future outcomes. A more formal analysis may be possible for other potential failure mechanisms. For example, PATH conducted analyses of the potential impacts of delays on dam breaching alternatives, taking into account the full range of other uncertainties, and thus was able to show that delaying breaching an additional 5 years (beyond the 3-year delay assumed as a breaching baseline) would reduce the weighted fraction of runs meeting the 24-year survival standard by 25% (Marmorek et al. 1998).

The draft Biological Opinion does include discussion of what it terms "best case" and "worst case" outcomes, which depend on assumptions about juvenile smolt survival and spawning escapements, and modified Dennis model assumptions about hatchery fish impacts. However, NMFS rejects some purportedly "worst case" analyses because NMFS deems them "beyond reason," even though extensive justification for these analyses had been developed PATH scientists.

Even so, assessing best and worst case scenarios can be a useful exercise, but the best available science indicates that there is considerably more to risk assessment than best case/worst case analyses (Burgman et al. 1993).

The prevailing practice for risk assessment in conservation biology is increasingly to use stochastic models instead of such “best case/worst case” exercises because such models allow analysis of the range of variability (Burgman et al. 1993, NRC 1995). For example, STUFA found that it was relatively straightforward to develop a variety of stochastic matrices for SRSSC based on the CRI Leslie matrices, because data were so readily available from PATH (Oosterhout et al. 2000). The advantage of stochasticizing these matrices was that it made it possible to incorporate at least some limited level of stochasticity into the analysis, instead of relying on analyses that set variability to zero and then analyze only best and worst case endpoints. Unexpected variability also can be incorporated into the analysis via sensitivity analysis or using the matrices to conduct simulated multi-factor experiments. Trends, density dependence, and other characteristics of real populations can be included, and results can be expressed as probability distributions.

As NMFS points out, developing stochastic simulations requires having some estimates of what the range of variability might be. Fortunately, as PATH and STUFA found, and the CRI acknowledges, there is an unusually complete set of data for the Snake River index spring and summer chinook populations. When stochastic simulations for these populations were carried out by STUFA, using CRI matrices but incorporating environmental, demographic, and observational uncertainties (albeit to limited extent), the equilibrium median population growth rates produced were much lower than the equivalent deterministic counterparts. The fact that stochastic models produce lower median population growth rates than do their deterministic counterparts is well-recognized. The more conservative results are biologically and mathematically expected (Burgman et al. 1993).

The STUFA results illustrating the tendency of stochastic models to produce less optimistic results than their deterministic counterparts help demonstrate how a limited “best/worst case” analysis is insufficient where more complete data is available as it is for SRSSC. These results also suggest that analyses for other populations where data is less available should be adjusted for the kinds of variability that simplistic best/worst case assessments omit. In addition, distributions from a stochastic Leslie matrix could be compared to distributions from the Dennis model, at least for SRSSC where data are available, so that the calibration of the two models could be completed. The fact that NMFS and CRI have not taken these steps, even where they could readily be taken, undermines the credibility and reliability of their work, certainly for SRSSC and possibly for other ESUs as well.

Risk assessment step 4: Develop a risk management strategy.

One difference between a rational process and one that is not rational is the degree to which decision-makers consider the mistakes they are inevitably making, and how to control, prevent, or mitigate potential impacts from the risk factors.

The RPA is in large part a plan to develop plans to detect further degradation, with the assumption that time will permit coming up with ways to control or mitigate risk mechanisms if needed. This plan assumes that (1) not enough is known now to be able to decide to take substantive action to control or mitigate for a number of important risks, (2) enough will be known in 5-8 years, and (3) actions to control or mitigate could be successfully implemented if needed.

Is it reasonable to accept these assumptions? What if they are wrong? A rational decision-maker would explicitly consider controlling, detecting, and mitigating for potential risks in case these assumptions are mistaken. For example, risk analysts often develop fault trees, which are essentially cause-and-effect diagrams that help clarify the chain of events that can cause a system to behave differently than expected. By examining the things that can go wrong, often with particular emphasis on cumulative effects or cascading events, decision-makers are better prepared to deal with the possibility that things won’t go as hoped. NMFS has evidently not carried out any such standard assessments. A risk assessment focused on risk factors and mechanisms, and what can go wrong, would clarify the implications and help decision-makers develop fail-safe and safe-fail mechanisms (Glickman and Gough 1990).

The approach of setting checkpoints is a necessary but insufficient start towards risk detection and management, but checkpoints are no more useful to risk management efforts than the quality of information relevant to decision-making that they are likely to provide. NMFS states that, at the 5 and 8 year checkpoints, "Resolution on critical uncertainties is necessary to assess progress" (NMFS 2000, p. 9-3); but because a Snake River spring/summer chinook salmon generation is approximately 5 years, there is little chance that changes in the key uncertainties could be detected at these checkpoints. As noted above, even a new overall long-term life-cycle survival rate using the Dennis model would require many more years. Analyses conducted by the CRI and PATH have shown that it will take decades for enough data to be collected to begin to resolve the most critical uncertainties (CRI 2000b, Peters et al. 2000). If resolution of the critical uncertainties is indeed necessary in 5 years, then the probability of achieving such unprecedented advances in scientific understanding should be included in the analysis as a critical uncertainty as well.

Although frequent checkpoints are important to adaptive management, it is also critical that experimental design methods make clear what can and cannot be detected at such checkpoints, and management is not "adaptive" unless mechanisms are in place to drive changes. Making a decision provisional on gathering more data is only justifiable if there is a mechanism in place that makes it feasible for such data to be gathered, and a mechanism that makes it feasible for management to respond.

Risk assessment step 5: Communicate.

Communication between stakeholders and decision-makers is an important aspect of ESA decision-making, and the requirements of the ESA and other environmental laws have helped support a great deal of effort on all sides to ensure that such communication is effective. However, this report is focused on the technical aspects of the decision process, rather than on the important aspects of the decision revolving around stakeholder buy-in, and how well decision-makers incorporate input from outsiders into the decision.

VIII. Sensitivity analysis

Sensitivity analysis is one of the most important elements of decision analysis. In decision analysis, the goals of sensitivity analysis are: to gain insights into the problem, to find a "simple and elegant structure that does justice to the problem," to validate the projections of likely outcomes and/or identify the appropriate limits on such projections, and to determine research priorities (von Winterfeldt and Edwards 1986, p. 387). This is not what CRI has used sensitivity analysis for.

Sensitivity analysis is most defensibly used during model development and validation, in order to gain insight into a model itself, because results can easily be biased by parameter definitions and analysis methods. In order to use sensitivity analysis to decide what courses of action should be taken, one has to show (1) that the models themselves represent the current, as well as future, real world adequately for the problem at hand; and (2) that the sensitivity analyses represent the way the models and the future world are likely to respond. Ordinarily, a model would have to undergo extensive validation, sensitivity analysis, and calibration before such conclusions could be justified.

Decision analysts rarely, if ever, use sensitivity analysis alone to drive decisions, partly because decision structuring allows a more thorough examination of all the elements of the decision. There is an important exception, however: *dominance* analysis is often one of the first steps in sensitivity analysis (Watson and Buede 1987). It is a useful, simple approach which examines the alternatives, uncertainties, and objectives, to determine whether, and under which conditions, a particular alternative might dominate the others. In other words, sometimes an alternative course of action is so robust to the elements and uncertainties in a decision that the choice is inevitable.

Even if no one alternative is absolutely dominant, sensitivity analysis can identify the circumstances under which the ranking of alternatives changes. This is one of the uses to which PATH put sensitivity analysis. As PATH showed, one reason dominance analysis is so useful is that it is directly helpful in boiling the decision down to its key elements.

There are several techniques for sensitivity analysis. The simplest, and potentially one of the most misleading, is to vary one parameter at a time by some percent, and see what percent effect it has on the response variable(s). There are two major problems with this approach: (1) the variables in complex systems are rarely completely independent of one another, and single-factor sensitivity analysis will ignore often-important compensatory or synergistic interactions such as negative and positive feedback processes; and (2) the results of a constant-percent method are determined entirely by the magnitude of the variables manipulated. Because of that, Caswell says the method "…provides no insight into sensitivity independently of those values" (Caswell 1989, p. 119).

For example, if mortalities for each life stage in a Leslie matrix are manipulated by a constant percent, as the CRI has done, then in a simple, linear model like the CRI's Leslie matrix, the result is foregone: the life stages with the highest mortalities will be shown to be the most "important." If it were that simple to design management plans, then managers would merely need to identify the stage with the highest mortality, and focus on that. But such simple solutions often miss the mark because the highest mortality may occur in a life stage where little can be done to improve survival. In such a situation, focusing efforts where mortality is highest may well lead to failure rather than success.

Consequently, in a rational decision process, decision-makers take into account what actions are likely to provide the greatest system response for the least effortful input. If they do not need to optimize a decision, they can focus on what it would take to at least achieve a minimum required response, such as avoiding jeopardy under the ESA. Either way, evaluating alternatives and uncertainties with an emphasis on efficiency and feasibility requires identifying the levers that can move the system as far as necessary with the least effort and risk. Such a common-sense determination requires analyzing what is feasible. The constant percent method of sensitivity analysis cannot help with that because its results are biased by the magnitudes of the parameters themselves.

Fortunately, there is a sensitivity analysis method[18] which has been shown to be useful, not only theoretically, but through numerous laboratory and field studies: elasticity analysis. Elasticity analysis is a method for examining system response to proposed management actions by standardizing the perturbations and responses, so that it produces an unbiased estimate of relative impact on the results. All the variables start from a level playing field, so the system response is not determined by the relative magnitude of the variables. Elasticity analysis is without question the standard method for conducting sensitivity analysis on matrix models in conservation biology (Caswell 1989, Caswell 2000, Crooks et al. 1998, Doak et al. 1994, Grant and Benton 2000, Heppell et al. 2000, Kroon et al. 2000, Ludwig 1999, Mills et al. 1999, Mode and Jacobson 1987, Wisdom et al. 2000). Because it helps minimize bias, it is also widely used in economics and other fields.

[18] The term "sensitivity analysis" here follows common usage to include any techniques for investigating how a model responds to perturbations. Strictly speaking, the mathematical method of formal sensitivity analysis is similar to elasticity analysis except for using absolute, rather than relative, perturbations — but to most people, "sensitivity analysis" has a more general meaning than that (Caswell 1989).

Although NMFS does not use it for the decision in the draft Biological Opinion, the CRI does report that results are significantly different when sensitivity analysis is conducted using the elasticity method. If this standard method were used to evaluate the sensitivity of the results to perturbations in the variables, attention might now be focused on adult migration survival rates, which have averaged around 50% since 1980 (CRI spreadsheet models, 1999-2000), taking into account Bonneville-to-Basin hydrosystem mortalities (40%), harvest (8%), and prespawning mortalities (10%). Instead, the focus has been turned to juvenile survival almost exclusively, simply by using the questionable constant percent method.

Elasticity analysis has its limitations as well, however, and one of them is that it is most accurate for small perturbations. Caswell (2000) points out that the most robust way to handle large perturbations, such as analyzing management options, is via numerical simulations. A particularly useful result of such an approach is that impacts of parameters on population growth estimates can be probabilistic (NRC 1995, Wisdom et al. 2000), a form of estimate that is particularly relevant to ESA assessments.

Numerical simulations can be dangerous too. For example, in risk assessment simulations, it is not uncommon to use Monte Carlo methods to vary each parameter over its expected range of variation and perhaps to include an element of uncertainty on top of that; then the response can be decomposed into the relative contributions of the component parts. This approach can be used to examine the relative impacts of uncertainties and variation on the model performance, particularly for retrospective models via ANOVA (Caswell 2000, Doak et al. 1994, Fahrig 1991, Oosterhout 1998). But what such an analysis shows is *not* what course of action should be taken (Caswell 2000). It shows only what the relative contributions of variability and uncertainty are to the modeled response.

Despite the recognized limitations of other sensitivity analysis methods, the fact remains that the method relied on by NMFS to evaluate management options is not even among the standard sensitivity analysis methods used for Leslie matrices, nor is it based on numerical simulations. Instead, NMFS relies on the simplistic constant percent method, applied only to a deterministic matrix for predictive, not validation, purposes. As noted, this method is misleading because the results are driven by the magnitude of the variables, rather than the relative sensitivity of the system to perturbations of those variables. More importantly, from a decision analysis point of view, NMFS inappropriately uses the results it obtains with this simple method to make management recommendations, without consideration of the feasibility of the parameter changes or the system responses assumed by the analysis. As a result, NMFS draws conclusions that cannot be supported by the analyses.

IX. Experimental design

The draft Biological Opinion emphasizes a perceived need to conduct experiments and otherwise gather more information in order to resolve uncertainties before a decision to breach the dams could be made. As noted above, a desire for more information is to be expected when important decisions have to be made in the face of uncertainty and political risk; but this hesitancy is only justified if the risks due to delay are likely to be more than offset by the increase in quality of the decision which will ultimately be made.

Delaying a decision in the face of uncertainty is appealing to decision-makers for two reasons: studying a problem some more and initiating new research agendas can seem like taking action, without actually implementing a strategy much different from the status quo; and there is always the chance that more information will indeed make the decision easier. Significant advances already made in information gathering and analysis over the past several years do not appear to have made this decision any easier, despite the fact that the SRSSC index populations are obviously in much worse straits now than they were when the 1995 decision launched PATH.

Under the ESA, the question here is thus whether the potential benefits of delaying a breaching decision outweigh the risks and costs of waiting, in order to conduct experiments to gather more information.

The most efficient way to carry out many kinds of experiments is not to carry them out at all, unless it is clear what they are needed for and what they can be expected to produce. In conjunction with the scientific method, analytical experiments via "Value of Information[19]" assessment is one of the most useful tools in decision analysis, and Value of Information assessments conducted in a structured decision framework have proven to be straightforward and useful aids to experimental design.

Value of Information assessment boils down to asking: if this uncertainty could be resolved, would the decision change, and if so, how (Hilborn and Walters 1992, von Winterfeldt and Edwards 1986)? This kind of analysis can save time and money that are often wasted in efforts to chase down information that will not have much effect on the choice of alternative, in the hope that more information would make the future clearer. It is also useful for prioritizing research efforts, helping narrow hypotheses down to testable assertions, and dealing with interactions among variables which can obfuscate the results.

If a logical decision structure has been developed, with the pertinent objectives, alternatives, and uncertainties clearly and appropriately defined, then Value of Information assessments should do much to resolve the question about which experiments really need to be done, if any, before a decision can be made. It also helps scientists design the experiments so that they do indeed provide the needed information.

Experimental design can be difficult enough even when a system is reasonably well-understood, and factors suspected of being important are controllable and measurable. But when a system is complex, with factors and relationships that are not well-understood, not controllable, and rarely directly measurable, experimental design is even more challenging. It can be easy to lose sight of what the questions are that truly need to be answered in order to make a decision, and whether a proposed experiment will actually provide answers that will be adequate to the task.

For example, as a formal decision analysis, PATH put primary emphasis on defining and analyzing hypotheses in order to understand how different uncertainties might affect the decision about how to avoid jeopardy to SRSSC, and thus what really needs to be further analyzed and

---

[19] A subset of Value of Information that requires a decision table is "Expected Value of Perfect Information," or EVPI (Hilborn and Walters 1992), which is essentially a dominance analysis.

tested, as indicated by the very name of the process: Plan for Analyzing and Testing Hypotheses. NMFS is not making use of the findings from this exceptionally thorough and rigorous approach, and instead is starting the experimental design process over, in a much more ad-hoc way.

In a politically risky decision process like this one, it is not surprising that NMFS is putting considerable effort into developing monitoring plans and identifying uncertainties and extensive, long term experimental programs such as those discussed in the All-H paper, and making little use of plans, programs, and analyses already available. Because the draft Biological Opinion is advocating a less-well defined process than PATH, it has failed to address two key experimental design questions, namely:

- Where is the burden of proof?

- What are the questions that need to be answered before a decision can be made, and is it feasible to get clear enough answers in the time allotted?

Taking these two points one at a time:

- **Where is the burden of proof?**

The most fundamental experimental design question in this case is where the burden of proof should lie. If a decision-maker is more risk averse for short-term economic or political threats, then the burden of proof will be on the species to show that economically or politically risky action is justified; if a decision-maker is risk averse for species extinction, then a more ecologically conservative course of action will be the default, and the burden of proof will be on the proposed course of action (e.g., the RPA) to demonstrate that it will not jeopardize the listed populations.

While the ESA may not demand the most ecologically conservative course of action, it does appear to require that the proposed course of action, not the species, bear the burden of risk. The NRC expresses concern "…that some current procedures may, perhaps inadvertently, bias decision-making in ways that are not intended under the act," (NRC 1995, p. 168), specifically these:

1. Statistical: using confidence interval definitions that disregard the probability of falsely accepting the null ("no effect") hypothesis; this puts the burden of proof on the need for more protection, or on scientists to demonstrate jeopardy — rather than on the need to show that more protection is *not* needed, or that an action will *not* cause jeopardy. This is in fact just what the CRI does in focusing on extinction probabilities as defined by confidence intervals in the modified Dennis model.

2. Cumulative effects/thresholds: the NRC (1995) argues that our technical ability to define thresholds and cumulative effects is limited. They recommend not simply assuming conservative but perhaps arbitrary thresholds, but instead recognizing that the consequences of each type of error be examined within a decision structure. That way, decision analysis "properly controls the risk of errors, from the point of view of species protection and from

the point of view of avoiding unneeded constraints" (p. 168). The draft Biological Opinion does not include the possibility of cumulative effects (temporally or spatially) or thresholds due to mechanisms such as depensation in the decision, despite the fact that they are sometimes mentioned in the text.

3. Listing decisions: the current practice of NMFS and USFWS is to not list unless the need is demonstrated, often requiring more information, which is particularly difficult to get when populations are already small. This approach places the burden of proof on the species, which the NRC points out is not consistent with the ESA. What the NRC advocates instead is not to simply reverse the burden of proof and confer protection on any proposed species by default, placing the burden of proof on showing that protection is not needed. Instead, they advocate examining the decision from the perspective of both kinds of errors (listing a species that does not need listing, or not listing a species that does need listing). If standard procedures are used to structure such a decision, consideration of both kinds of errors occurs by default (Hilborn 1997). The NMFS analysis in the draft BiOp resolves this problem by ignoring it.

4. Reducing asymmetry of risk for listed species: The NRC argues that the usual approach is to pose the null hypothesis of no harm, and set a low probability (e.g., 5%) of falsely concluding no harm — implicitly or explicitly, depending on how well structured the decision process is. This approach "…is more likely to deny needed protection than to afford unneeded protection" (p. 169). What the NRC advocates is the "precautionary principle," which they point out has been endorsed in several international legal documents. Interestingly, an example precautionary approach they cite is NMFS' decision to list Snake River sockeye, even in the face of uncertainty about its status as a distinct species.

It is not clear how NMFS might justify not taking the appropriately risk-averse approach, when it is clearly stated in the draft Biological Opinion (NMFS 1995) that:

> *In the absence of a recovery plan, the uncertainty of whether an action avoids jeopardy or adverse modification of critical habitat is greater. Therefore, an action must avoid or offset adverse effects to the listed species to a greater extent than could likely be determined with the benefit of recovery planning. This extra effort is necessary to meet the statutory requirement that an Action Agency 'ensure,' in the face of uncertainty, that its action is not likely to jeopardize the species or adversely modify critical habitat (p. 1-9).*

The fact that NMFS has chosen the RPA, despite the fact that the RPA does not meet the jeopardy standards for several of the Snake River index populations (even under what the draft Biological Opinion acknowledges as optimistic assumptions), indicates that NMFS has chosen to place the burden of proof on the listed populations. If this is so, and it appears to be the case based on the analysis in the draft Biological Opinion, then the experimental design incorporated into the draft Biological Opinion would appear to proceed from fundamentally the wrong premise, i.e., one that requires firm proof that protection is required and discounts or ignores the risk of failing to take protective steps.

- **What are the questions that need to be answered before a decision can be made, and is it feasible to get clear enough answers in the time allotted?**

A second experimental design issue is what lack of information is holding up a decision, and whether it is even possible to get that lack of information resolved well enough to make the decision easier to implement or justify than it is now. Before embarking on an expensive and risky course of action based on an assumption that more data will help mitigate and avoid risks, this is a question that a rational decision process should confront.

The list of questions that scientists would like to answer is long and somewhat mutable, as indicated by the fact that a substantial portion of the draft Biological Opinion (and the All-H paper) consists of a proposal to develop another plan for resolving uncertainties. Nonetheless, there is little dispute that extra mortality believed to be due to the hydrosystem is a key uncertainty, and one aspect of that issue is the relative magnitude of delayed mortality for transported juveniles versus non-transported fish.

It may be useful here to give an example of what might be involved in resolving the issue of relative survival rates of transported and in-river juvenile salmon. The parameter representing this, *D*, is defined as (Bouwes et al. 1999):

> *…the differential survival rate of transported fish relative to fish that migrate in-river, as measured from BON tailrace to adult returning to Lower Granite Dam (LGR). A 'D' equal to one indicates that there is no difference in survival rate (after hydrosystem passage), while a 'D' less than one indicates that transported fish die at a higher rate after release, than fish that have migrated through the hydrosystem (p. 3).*

D is not a parameter that can be measured. *D* is a calculated estimate that requires knowing many things. For example: how many smolts are picked up by barge at Lower Granite, Little Goose, and Lower Monumental dams during at least the 2-month migration peak? How many smolts travel through the hydrosystem before they are picked up by barges, at each pickup point? How many smolts travel all the way through the hydrosystem on their own? And what are the survival rates to adulthood for each of those groups? Other factors that could affect the results are inter- or intra-annual variation in Fish Guidance Efficiencies, spill effectiveness, and collection system survival; inland and ocean climate patterns; and adult migration survival rates. Schreck et al. (in prep.) found that survival rates drop during the April-to-June migration peak, and so that might have to be taken into account as well. And finally, there is considerable interest in hatchery vs. wild fish survival rates.

The best way to estimate these fractions and survival rates is currently through PIT-tag tracking. Hatchery fish can be PIT-tagged at the hatchery, but wild fish have to be intercepted and tagged at some collection point, both as juveniles and then again as returning adults. There are not many wild fish left, and tagging a large enough sample of non-hatchery fish is a problem.

How many smolts would need to be tagged? Sample size depends on experimental design requirements, such as how exact a prediction is needed and how confident decision-makers need to be, as well as what the true values are and how different they are from the hypothesized values. Peters et al. (2000) provide some examples to set the range of sample sizes that might be required. In an example where they assume that 1975-1999 conditions are representative, they found that if the true *D* is high (0.7), and survivals (smolt-to-adult returns or SARs) are very low (0.25%), in order to be able to tell if the true *D* is less than 0.65, for just one year, it would require tagging 900,000 barged smolts and 1.2 to 12 million in-river smolts (Peters et al. 2000, Table ES-5, p. ES-11). If instead, the true *D* is much higher than anyone thinks it is (1.0), and SARs are the minimum that some people think they need to be for survival (2%), then the contrast would be larger between the hypothesized and real values, and the difference would be easier to detect. In that case, it would require tagging only 3,350 transported smolts and 6,700 to 67,000 in-river smolts to tell if the true *D* in a single year is less than 0.65. To get some idea of how these sample size relates to the population, in recent years, there have been around 35,000 to 50,000 smolts total migrating out from the 7 Snake River index streams, which make up something on the order of about 10% of the Snake River population.

Determining what *D* is in one year does not resolve the problem, however. The question is what *D* can be expected to be over time, and that requires developing confidence intervals and thus multiple years and multiple extensive tagging studies. Peters et al. also analyzed how many years of data would be required to develop various confidence levels (Figure ES-8, p. ES-11). Depending on experimental design requirements, they found that in order to develop a 90% confidence that true *D* is greater than 0.65, if the measured *D*'s were less than about 0.8, even 80 years of data might not be enough. Further complicating the issue is the fact that as noted above, *D* is not a measured quantity, but rather a derived variable, and the method for calculating it remains in dispute.

This is not an argument that further study of *D* is unnecessary. The issue is that SRSSC and other listed salmon populations are already in steep decline; many populations have already been lost and the probability of extinction is high for others. PATH and STUFA have concluded that although experiments like the ones NMFS and CRI have proposed are important, waiting for such experiments to be designed, carried out and analyzed, means accepting a high probability of extinction in the meantime for many of the Columbia and Snake River salmon populations, and a low probability that the uncertainties would ever be resolved anyway.

X. Conclusions

This paper has analyzed the decision process evidenced by the draft Biological Opinion regarding Snake River spring and summer chinook salmon populations listed as threatened under the ESA. The widely acknowledged standard for logical, rational decision making under conditions of uncertainty is the science of decision analysis (Clemen 1996, Keeney 1992, Keeney and Raiffa 1976, NRC 1995, von Winterfeldt and Edwards 1986), which thus provides the appropriate framework from which to evaluate the NMFS decision process.

An extensive, rigorous, peer-reviewed decision analysis has already been conducted for SRSSC (i.e., PATH), but NMFS has chosen not to accept its conclusions. Rather than accept the existing decision analysis, or conduct some other decision analysis within a structured framework, as recommended by the NRC (1995), NMFS is basing the draft Biological Opinion on a collection of unvalidated analytical models, professional judgments, and a plan to fill in the blanks sometime in the future. The problem is that this collection of models, opinions, and proposals does not provide an analysis that supports the conclusion drawn.

By any definition, rational decision making requires clearly defined objectives and representative metrics, unambiguously defined alternatives, identification of uncertainties and their incorporation into the analysis, evaluation of the potential impacts of errors and risks to implementation and success of each alternative, and the use of tools such as Value of Information assessments and sensitivity analysis to validate the decision structure, develop an adaptive management framework, prioritize research needs, and develop planned experiments. This approach is not only good decision science, it is simple common sense.

The decision process evidenced by the draft Biological Opinion is inadequate on every count:

- NMFS has not explicitly analyzed the decision within the framework of clearly defined objectives, as required by the Endangered Species Act and recommended by the National Research Council, because the metrics are not representative of the jeopardy standards as stated, and the analytical methods are not consistent with the best available science for estimating such metrics.

- The preferred alternative, namely the RPA and offsite mitigation, has yet to be clearly defined but is assumed to be adequate to avoid jeopardy anyway.

- Uncertainties, variabilities, and errors have not been incorporated directly into the analysis of the RPA, and the ultimate conclusions do not take them into account. Even though often acknowledged in the text, they are not incorporated into the quantitative analyses, nor are they incorporated into the conclusions.

- The models are simplistic, inappropriately applied, statistically problematic, and are not consistent with prevailing practice.

- A true risk assessment, which would incorporate into the analysis uncertainties and factors that can affect achievement of the objectives, has not been incorporated into the decision. Instead, assumptions and simplifications on which the decision is based are systematically optimistic.

- The risk management approach, which relies on near-term checkpoints, fails to acknowledge that the data on which these checkpoints are to be based cannot possibly be adequate in the time frame of a single salmon generation.

As a result of these decision process failures, the linkages between the pieces of the purported chain of analyses fail to provide a continuous, coherent pathway by which the analyses can be shown to support the conclusions drawn. Neither the decision process nor the available science supports a decision that is so vaguely defined and poorly analyzed. Largely because of almost exclusive reliance on deterministic exponential population growth projections into the future, the overall approach is highly optimistic, and fails to make good use of a great deal of available research and analysis, particularly for the Snake River spring and summer chinook ESU. Although this discussion has focused on Snake River spring and summer chinook, the decision process failures identified apply to all 12 Columbia/Snake River ESUs.

XI. References Cited

Akcakaya, H. R., S. Ferson, M. A. Burgman, D. A. Keith, G. M. Mace, and C. R. Todd. 2000. Making consistent IUCN Classifications under uncertainty. *Conservation Biology* 14: 1001-1013.

Allee, W. C., A. E. Emerson, O. Park, T. Park, and K. P. Schmidt. 1949. *Principles of Animal Ecology*. Saunders, Philadelphia.

Botsford, L. W., and J. G. Brittnacher. 1998. Viability of Sacramento River winter-run chinook salmon. *Conservation Biology* 12(1): 65-79.

Bouwes, N., H. A. Schaller, P. E. Budy, C. E. Petrosky, R. Kiefer, P.Wilson, O. P. Langness, E. Weber, and E. Tinus. 1999. An Analysis of Differential Delayed Mortality Experienced by Stream-type Chinook Salmon of the Snake River: A response by State, Tribal, and USFWS technical staff to the 'D' analyses and discussion in the Anadromous Fish Appendix to the U.S. Army Corps of Engineers' Lower Snake River Juvenile Salmonid Migration Feasibility Study.

Burgman, M. A., S. Ferson, and H. R. Akcakaya. 1993. *Risk Assessment in Conservation Biology*. Chapman and Hall, UK.

Byers, R. E., and R. I. C. Hansell. 1992. Stability-like properties of population models. *Theoretical Population Biology* 42: 10-34.

Caswell, H. 1989. *Matrix population models*. Sinauer Associates, Inc. Publishers, Sunderland, MA.

Caswell, H. 2000. Prospective and retrospective perturbation analyses: their roles in conservation biology. *Ecology* 81: 619-627.

Clemen, R. T. 1996. *Making Hard Decisions: An Introduction to Decision Analysis*. Duxbury Press, Wadsworth Publishing Co., Belmont, CA.

CRI (Cumulative Risk Initiative). 2000a (7-April 2000 draft). A standardized quantitative analysis of risks faced by salmonids in the Columbia River Basin. Northwest Fisheries Science Center NMFS - NOAA, Seattle, WA.

CRI (Cumulative Risk Initiative). 2000b (17-July 2000 draft). A standardized quantitative analysis of risks faced by salmonids in the Columbia River Basin. Northwest Fisheries Science Center NMFS - NOAA, Seattle, WA.

Crooks, K. R., M. A. Sanjayan, and D. F. Doak. 1998. New insights on cheetah conservation through demographic modeling. *Conservation Biology* 12(4): 889-895.

DeAngelis, D. L., and W. Post, M. 1991. Positive feedback and ecosystem organization. Pages 155-178 in M. Higashi and T. P. Burns, eds. *Theoretical studies of ecosystems*. Cambridge University Press, Cambridge.

Dennis, B., P. L. Munholland, and J. M. Scott. 1991. Estimation of growth and extinction parameters for endangered species. *Ecological monographs* 61(2): 115-143.

Derby, S. L., and R. L. Keeney. 1990. Risk analysis: understanding "How safe is safe enough?". Pages 43-50 in T. S. Glickman and M. Gough, eds. *Readings in Risk*. Resources for the Future, Washington, DC.

Doak, D., P. Kareiva, and B. Klepetka. 1994. Modeling population viability for the desert tortoise in the western Mojave desert. *Ecological Applications* 4(3): 446-460.

Fahrig, L. 1991. Simulation methods for developing general landscape-level hypotheses of single-species dynamics. Pages 417-442 in M. G. Turner and R. H. Gardner, eds. *Quantitative Methods in Landscape Ecology: the Analysis and Interpretation of Landscape Heterogeneity*. Springer-Verlag, New York.

Federal Caucus (Army Corps of Engineers, Bonneville Power Administration, Bureau of Indian Affairs, Bureau of Land Management, Bureau of Reclamation, Environmental Protection Agency, Fish and Wildlife Service, Forest Service, and National Marine Fisheries Service). 2000. *Conservation of Columbia Basin Fish* Draft Basin-wide Salmon Recovery Strategy (Update of the All- H Paper). Federal Caucus, Portland.

Glickman, T. S., and M. Gough, eds. 1990. *Readings in Risk*. Resources for the Future, Washington DC.

Grant, A., and T. G. Benton. 2000. Elasticity analysis for density-dependent populations in stochastic environments. *Ecology* 81(3): 680-693.

Heppell, S. S., H. Caswell, and L. B. Crowder. 2000. Life histories and elasticity patterns: perturbation analysis for species with minimal demographic data. *Ecology* 81(3): 654-665.

Higgins, K. 1999. The impact of age- and space- and stochastic-structure on the extinction probability of a chinook salmon metapopulation. *Natural resource modeling* 12(1): 65-108.

Hilborn, R., and C. J. Walters. 1992. *Quantitative fisheries stock assessment*. Chapman and Hall, New York.

Hilborn, R., and M. Mangel. 1997. *The ecological detective: confronting models with data*. Princeton University Press, Princeton.

Holling, C. S. 1976. Resilience and stability of ecosystems. Pages 73-92. *Evolution and Consciousness*. Addison-Wesley Publishing Co., Reading, MA.

Holling, C. S., ed. 1978. *Adaptive environmental assessment and management*. John Wiley and Sons, Chichester.

Holmes, E. E. 2000 in prep. Estimating risks in declining populations with poor data.

Howard, R. A. 1968. The foundations of decision analysis. *IEEE Transactions on systems science and cybernetics* SSC-4(3): 211-219.

Howard, R. A. 1989. Decision analysis: applied decision theory. Pages 97-113 in R. A. Howard and J. E. Matheson, eds. *The principles and applications of decision analysis*. Strategic Decisions Group, Menlo Park, CA.

ISAB (Independent Scientific Advisory Board). 1999. Review of the National Marine Fisheries Service draft Cumulative risk Analysis addendum. ISAB; report prepared for NWPPC and NMFS, Portland, OR.

Janis, I. L. 1972. *Victims of groupthink: a psychological study of foreign-policy decisions and fiascos*. Houghton Miflin, Boston.

Kareiva, P., M. Marvier, M. McClure, P. McElhany, M. Ruckelshaus, B. Sanderson, and R. Waples. 1999 (draft). An introduction to NMFS decision-support science for ESA decision making, with examples. National Marine Fisheries Service, Northwest Fisheries Science Center, Seattle, WA.

Keeney, R. 1992. *Value-Focused Thinking*. Harvard University Press, London.

Keeney, R. L., and H. Raiffa. 1976. *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*. John Wiley and Sons, New York.

Kroon, H. D., J. V. Groenendael, and J. Ehrlen. 2000. Elasticities: a review of methods and model limitations. *Ecology* 81(3): 607-618.

Ludwig, D. 1999. Is it meaningful to estimate a probability of extinction? *Ecology* 80(1): 298-310.

Ludwig, D., and C. J. Walters. 1985. Are Age-Structured Models Appropriate for Catch-Effort Data? *Can. j. Fish. Aquat. Sci.* 42: 1066-1072.

Marmorek, D. R., J. J. Anderson, L. Bashan, D. Bouillon, T. Cooney, R. Deriso, P. Dygert, L. Garrett, A. Giorgi, O. P. Langness, D. Lee, C. McConnaha, I. Parnell, C. M. Paulsen, C. Peters, C. E. Petrosky, C. Pinney, H. A. Schaller, C. Toole, E. Weber, P. Wilson, and R. W. Zabel. 1996. Plan for Analyzing and Testing Hypotheses (PATH): Final report on retrospective analyses for fiscal year 1996. ESSA Technologies, Vancouver, BC.

Marmorek, D. R., C. N. Peters, and I. Parnell. 1998. PATH final report for fiscal year 1998. ESSA Technologies Ltd., Vancouver, BC.

McElhany, P., M. Ruckelshaus, M. J. Ford, T. Wainwright, and E. Bjorkstedt. 2000. Viable salmonid populations and the recovery of evolutionarily significant units. National Marine Fisheries Service, Seattle, WA.

Mills, L. S., D. F. Doak, and M. J. Wisdom. 1999. Reliability of conservation actions based on elasticity analysis of matrix models. *Conservation Biology* 13(4): 815-829.

Mode, C. J., and M. E. Jacobson. 1987. A study of the impact of environmental stochasticity on extinction probabilities by Monte Carlo integration. *Mathematical BioSciences* 83: 105-125.

Mundy, P. R. 1999. Status and expected time to extinction for Snake river spring and summer chinook stocks: the doomsday clock and salmon recovery index models applied to the Snake river Basin. Fisheries and Aquatic Sciences, Lake Oswego, OR.

Musick, J. A. 1999. Criteria to define extinction risk in marine fishes: the American Fisheries Society Initiative. *Fisheries* 24 (12): 6-14.

Nickelson, T. E., and P. W. Lawson. 1998. Population viability of coho salmon, Oncorhynchus kisutch, in Oregon coastal basins: application of a habitat-based life cycle model. *Canadian Journal of Fisheries and Aquatic Sciences* 55: 2383-2392.

NMFS (National Marine Fisheries Service). 2000. Endangered Species Act - Section 7 Consultation, DRAFT Biological Opinion: Operation of the Federal Columbia River Power System Including the Juvenile Fish Transportation Program and the Bureau of Reclamation's 31 Projects, Including the Entire Columbia Basin Project. Consultation conducted by NMFS; agencies are: U.S. Army Corps of Engineers Bonneville Power Administration, Bureau of Reclamation, National Marine Fisheries Service, Seattle.

Noss, R.F., M.A. O'Connell, and D.P. Murphy. 1997. *The Science of Conservation Planning*. Island Press. Washington, D.C.

NRC (National Research Council). 1995. *Science and the Endangered Species Act*. National Academy Press, Washington, D.C.

Oosterhout, G. 1998. *PasRAS*: a stochastic simulation of chinook and sockeye life histories. Report available from Portland General Electric, Portland, OR.

Oosterhout, G., P. Wilson, C. Petrosky, P. Budy, N. Bouwes, E. Weber, E. Tinus, G. Thiede, and H. Schaller. 2000. A technical review of the National Marine Fisheries Service Leslie matrix model of Snake River spring and summer chinook populations. STUFA: State, Tribal, and U.S. Fisheries Agencies, Portland, OR.

PATH Scientific Review Panel. 1998. Conclusions and recommendations from the PATH Weight of Evidence Workshop, Vancouver BC.

Peters, C. N., D. Marmorek, J. Anderson, N. Bouwes, P. Budy, T. Cooney, R. Deriso, A. Giorgi, R. Hinrichsen, B. Muir, C. Paulsen, C. Petrosky, S. Smith, E. Weber, and R. Zabel. 2000. PATH: Preliminary Evaluation of the Learning Opportunities and Biological Consequences of Monitoring and Experimental Management Actions. Calvin N. Peters and David Marmorek, (compilers / editors). ESSA Technologies Ltd. 1765 West 8th Avenue, Suite 300 Vancouver, BC V6J 5C6.

Punt, A. E., and R. Hilborn. 1997. Fisheries stock assessment and decision analysis: the Bayesian approach. *Reviews in Fish Biology and fisheries* 7: 35-63.

Ratner, S., R. Lande, and B. B. Roper. 1997. Population viability analysis of spring chinook salmon in the South Umpqua River, Oregon. *Conservation Biology* 11(4): 879-889.

Russo, J. E., and P. J. H. Shoemaker. 1989. *Decision traps: the ten barriers to brilliant decision-making and how to overcome them*. Fireside Books, New York.

Schreck, C. B., C. S. T. S. Darren Lerner, and Larry Davis, G. Oosterhout, and J. B. Congleton. In prep. (contact: Carl Schreck, Oregon Cooperative Fishery Research Unit, Oregon State University, Corvallis) Interaction of stress, pathogens and development on behavior of teleosts.

Sonnenberg, F. A., and R. Beck. 1993. Markov models in medical decision making: a practical guide. *Medical Decision Making* 13(4): 322-173.

Thompson, G. G. 1991. Determining minimum viable populations under the Endangered Species Act. National Marine Fisheries Service, Alaska Fisheries Science Center, Seattle.

von Winterfeldt, D., and W. Edwards. 1986. *Decision Analysis and Behavioral Research*. Cambridge University Press, Cambridge.

Vose, D. 2000. *Risk analysis: a quantitative guide*. John Wiley and Sons, Chichester.

Wainwright, T., and R. Waples. 1998. Prioritizing Pacific salmon stocks for conservation: response to Allendorf et al. *Conservation Biology* 12: 1144-1147.

Watson, S. R., and D. M. Buede. 1987. *Decision synthesis: the principles and practice of decision analysis*. Cambridge University Press, Cambridge.

Wisdom, M. J., L. S. Mills, and D. F. Doak. 2000. Life stage simulation analysis: estimating vital-rate effects on population growth for conservation. *Ecology* 81: 628-641.